The Open University

M303

Further pure mathematics

Book C

Numbers and rings

This publication forms part of the Open University module M303 *Further pure mathematics*. Details of this and other Open University modules can be obtained from Student Recruitment, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)300 303 5303; email general-enquiries@open.ac.uk).

Alternatively, you may visit the Open University website at www.open.ac.uk where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2014. Second edition 2016.

Copyright © 2014, 2016 The Open University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted or utilised in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission from the publisher or a licence from the Copyright Licensing Agency Ltd. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS (website www.cla.co.uk).

Open University materials may also be made available in electronic formats for use by students of the University. All rights, including copyright and related rights and database rights, in electronic materials and their contents are owned by or licensed to The Open University, or otherwise used by The Open University as permitted by applicable law.

In using electronic materials and their contents you agree that your use will be solely for the purposes of following an Open University course of study or otherwise as licensed by The Open University or its assigns.

Except as permitted above you undertake not to copy, store in any medium (including electronic storage or use in a website), distribute, transmit or retransmit, broadcast, modify or show in public such electronic materials in whole or in part without the prior written consent of The Open University or in accordance with the Copyright, Designs and Patents Act 1988.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed in the United Kingdom by Halstan & Co. Ltd, Amersham, Bucks.

ISBN 978 1 4730 2038 2

2.1

Contents

Chapter 9 Multiplicative functions **1**

Introduction **3**

1 Multiplicative functions **4**
1.1 The τ and σ functions 4
1.2 Multiplicative functions 7

2 Perfect numbers **12**
2.1 Even perfect numbers 12
2.2 Mersenne primes 14
2.3 Some results 17

3 Euler's ϕ-function **21**
3.1 Euler's Theorem 21
3.2 The order of an integer to a composite modulus 25
3.3 Applications of Euler's Theorem 26

4 Properties of Euler's ϕ-function **31**
4.1 The multiplicativity of Euler's ϕ-function 31
4.2 The formula for $\phi(n)$ 33

5 Primitive roots **37**
5.1 Primitive roots generalised 37
5.2 Moduli admitting primitive roots 41

Learning outcomes **44**

Solutions to exercises **45**

Chapter 10 Quadratic reciprocity **59**

Introduction **61**

1 Euler's Criterion **62**
1.1 Quadratic residues 62
1.2 Euler's Criterion 66
1.3 Uses of Euler's Criterion 68

2 The Legendre symbol **71**
2.1 Properties of the Legendre symbol 71
2.2 Applications using the Legendre symbol 73

3 Gauss's Lemma **75**
3.1 Proof of Gauss's Lemma 75
3.2 Applications of Gauss's Lemma 78
3.3 More cases of Dirichlet's Theorem 81

4 The Law of Quadratic Reciprocity **83**
4.1 Understanding the Law 83
4.2 Proof of LQR 86
4.3 Applications of LQR 90

Learning outcomes **94**

Solutions to exercises **95**

Chapter 11 Rings and polynomials **109**

Introduction **111**

1 Rings **111**
1.1 Subrings 117
1.2 Zero divisors and units 119
1.3 Fields 122

2 Polynomials over fields **125**
2.1 Polynomial rings 125
2.2 The degree of a polynomial 127
2.3 Basic properties of polynomial rings over fields 129

3 Divisibility of polynomials **131**
3.1 Polynomial division 131
3.2 Highest common factors 135
3.3 The Euclidean Algorithm 137
3.4 Least common multiples 139

4 Factorising polynomials **142**
4.1 Roots and irreducible polynomials 142
4.2 Polynomials in $\mathbb{Z}[x]$ and rational roots 145
4.3 Gauss's Lemma 147
4.4 Further tests for irreducibility 151

Learning outcomes **157**

Solutions to exercises **158**

Chapter 12 Fermat's Last Theorem and unique factorisation **171**

Introduction **173**

1 Fermat's Last Theorem and Diophantine equations **173**
1.1 Pythagorean triples 174
1.2 The equation $x^4 + y^4 = z^4$ 180
1.3 Related Diophantine equations 183

2 Integral domains **185**
2.1 Units and associates 186
2.2 Factors, irreducibles, primes and integral domains 187
2.3 Norms 192

3 Euclidean domains **194**
3.1 Rings with a division algorithm 194
3.2 Highest common factors 198
3.3 Prime numbers expressed as the sum of two squares 202

4 Unique factorisation domains **205**
4.1 Prime and irreducible in unique factorisation domains 205
4.2 UFDs and Euclidean domains 208
4.3 Sums of two squares, revisited 212
4.4 More Diophantine equations 215

Learning outcomes **217**

Solutions to exercises **218**

Index **229**

Chapter 9

Multiplicative functions

Introduction

Welcome to Book C *Numbers and rings*. This book begins by building on what you learned in Book A, before explaining some generalisations of number theory in the study of rings.

From the time of Fermat's announcement of Fermat's Little Theorem (FLT), 120 years passed before a proof was published. That proof was given by Euler (1707–1783) who, some 24 years later, went on to prove a generalisation of FLT that caters for the case of the modulus being composite. In doing so, he introduced a function that we nowadays refer to as Euler's *ϕ-function*; it is a function whose importance in number theory cannot be overestimated. The ϕ-function, its properties and the generalisation of FLT form the core of this chapter.

Recall that Euler is pronounced *oiler*.

ϕ is the Greek letter 'phi'.

In this chapter we are much concerned with *number-theoretic functions*, that is, integer-valued functions defined on the set of natural numbers. When a number-theoretic function f has the property that $f(mn) = f(m)f(n)$ whenever $\text{hcf}(m, n) = 1$, it is called a *multiplicative function*. Multiplicative functions will be found to have the desirable property that their values everywhere are determined by their values at prime powers. We will meet numerous examples of multiplicative functions in this chapter, the ϕ-function being one of them. However, we begin by investigating two other classical multiplicative functions: the τ function, which counts the number of factors of a natural number, and which you have already encountered in Book A, Chapter 2; and the σ function, which adds up the factors of a natural number.

Recall that τ is the Greek letter 'tau', and σ is the Greek letter 'sigma'.

Linked to the study of the σ function is the following special case. The number 6 has the interesting property that it is equal to the sum of all its factors excluding itself:

$$6 = 1 + 2 + 3.$$

Numbers with this property are called *perfect numbers* and, as reported in this chapter, not many have been found. The search for perfect numbers has a fascinating history, having occupied mathematical minds since the time of the early Greeks – and it is still pursued actively today. When investigating perfect numbers you will meet the so-called *Mersenne primes*: a family of rare primes, each of which yields a perfect number.

In Book A, Chapter 4 we touched upon the ideas of order and primitive roots when working to some prime modulus. With the benefit of the ϕ-function we look again at those notions, generalising and extending the earlier work to cater for a composite modulus.

Throughout this chapter and the next, one name will keep appearing: that of Euler. You can learn more about his life in the *History Reader* for this topic. As far as number theory is concerned, we have already mentioned that it was Euler who came up with the first published proof of FLT. You will see more of his contributions in this chapter.

Chapter guide

This chapter consists of five sections. Section 1 introduces multiplicative functions, and forms the foundation for the rest of the chapter. Section 2 explores perfect numbers and related concepts, and is not required by later sections. The final three sections focus on Euler's ϕ-function, with Section 3 forming the basis for the last two sections. The final section is shorter, and does not rely on content from Section 4.

1 Multiplicative functions

This section begins by looking at two multiplicative functions, and then uses properties of these to infer some more general results.

1.1 The τ and σ functions

In Book A, Chapter 2 we met the τ function. For convenience we recall the definition here.

Definition 1.1 *The τ function*

For any integer $n \geq 1$, $\tau(n)$ is defined to be the number of distinct factors of n, including 1 and n.

The τ function is an example of an integer-valued function that has the set of natural numbers, $\mathbb{N}$, as its domain. We refer to such functions $f : \mathbb{N} \to \mathbb{Z}$ as being **number-theoretic functions**. Although a number-theoretic function is permitted to take any integer value, in all the examples in this chapter the values taken will be positive integers or 0.

Number-theoretic functions are also called **arithmetic functions**.

Closely related to the τ function is a second number-theoretic function, denoted by σ.

Definition 1.2 *The σ function*

For any integer $n \geq 1$, $\sigma(n)$ is defined to be the sum of the distinct factors of n, including 1 and n.

So, for example, as the factors of 20 are 1, 2, 4, 5, 10 and 20 we have

$$\sigma(20) = 1 + 2 + 4 + 5 + 10 + 20 = 42.$$

Exercise 1.1

Complete the following table.

n	1	2	3	4	5	6	7	8	9	10
$\tau(n)$										
$\sigma(n)$										

Before examining these two functions in more detail, let us introduce some notation that will be of assistance in a number of situations. The Greek letter 'capital sigma', Σ, is customarily used to denote summation. We extend that notation and write

$$\sum_{d|n} f(d)$$

to denote the sum of terms $f(d)$ taken over all positive factors d of n. For example,

$$\sum_{d|10} f(d) = f(1) + f(2) + f(5) + f(10).$$

With this notation we observe the following.

Factor sum formulas for τ and σ

$$\tau(n) = \sum_{d|n} 1 \quad \text{and} \quad \sigma(n) = \sum_{d|n} d$$

In addition to summation, we are concerned in this chapter with products, for which we have a corresponding notation in which sigma is replaced by the Greek letter 'capital pi', Π. The following examples illustrate typical uses.

$$\prod_{1\le d\le 4} f(d) = f(1)f(2)f(3)f(4)$$

$$\prod_{\substack{p|28\\ p \text{ prime}}} f(p) = f(2)f(7)$$

In the second of these examples, the product is taken over all prime factors of 28.

Exercise 1.2

(a) Determine $\sum_{d|8} \sigma(d)$ and $\prod_{d|8} \tau(d)$.

(b) (i) Show that

$$\sum_{d|n} f(d) = \sum_{d|n} f\left(\frac{n}{d}\right).$$

Hint: if d is a factor of n then $\frac{n}{d}$ is also a factor of n.

(ii) Deduce from this formula that

$$\sum_{d|n} \frac{1}{d} = \frac{\sigma(n)}{n}.$$

If n is composite then $\tau(n) > 2$ and $\sigma(n) > n + 1$.

It is a simple matter to determine the values taken by each of the τ and σ functions at primes. If p is a prime then, as the only factors of p are itself and 1, $\tau(p) = 2$ and $\sigma(p) = p + 1$. Indeed, an integer $n > 1$ is prime if, and only if, $\tau(n) = 2$, or equivalently, if, and only if, $\sigma(n) = n + 1$, for each of these statements amounts to the fact that the only factors of n are itself and 1.

We can readily go a stage further and look at powers of the prime p. As the factors of p^k are 1, p, p^2, ..., p^k, we have

$$\tau\left(p^k\right) = k + 1 \quad \text{and} \quad \sigma\left(p^k\right) = 1 + p + p^2 + \cdots + p^k = \frac{p^{k+1} - 1}{p - 1}.$$

The second expression for $\sigma(p^k)$ utilises the formula for the sum of a geometric series obtained in Exercise 3.2 of Book A, Chapter 1.

In Book A, Chapter 2 we completed the task for the τ function, obtaining the following formula for the value of $\tau(n)$ at any natural number n, where n is given in its prime decomposition:

$$\tau\left(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\right) = (k_1 + 1)(k_2 + 1) \cdots (k_r + 1).$$

We may refer to this as the prime decomposition formula for τ.

But notice that as $\tau\left(p_i^{k_i}\right) = (k_i + 1)$, we obtain the following formula:

$$\tau\left(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\right) = \tau\left(p_1^{k_1}\right) \tau\left(p_2^{k_2}\right) \cdots \tau\left(p_r^{k_r}\right).$$

This shows that the value of the τ function at any integer n is completely determined by its values at the prime power factors of n. In other words, we know all about τ once we know its value at a general prime power. Functions with this property command considerable interest in number theory.

1.2 Multiplicative functions

Definition 1.3 *Multiplicative functions*

A number-theoretic function is said to be **multiplicative** if, for any integer with prime decomposition $p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$,

$$f\left(p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}\right) = f\left(p_1^{k_1}\right)f\left(p_2^{k_2}\right)\cdots f\left(p_r^{k_r}\right).$$

So τ has provided us with our first example of a multiplicative function, and it will come as no surprise to learn that the σ function provides a second example. Before establishing that fact, let us practise making use of the multiplicative nature of such functions.

Exercise 1.3

In this exercise you may assume the result that τ and σ are multiplicative functions.

(a) Determine $\tau(360)$ and $\sigma(360)$.

(b) If $p > 3$ is prime, determine $\tau(2p^3)$, $\sigma(3p)$ and $\sigma(6p^2)$.

The ability to calculate values of multiplicative functions at composite numbers in the way illustrated in Exercise 1.3 will prove to be very useful. However, there is an alternative way of viewing the definition of a multiplicative function that will simplify matters in some situations.

Proposition 1.4 *Equivalent definition of multiplicative functions*

The function f is a multiplicative function if, and only if, for every pair, m and n, of coprime natural numbers,

$$f(mn) = f(m)f(n).$$

Proof Suppose first that f is multiplicative, and let $m = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$ and $n = q_1^{l_1}q_2^{l_2}\cdots q_s^{l_s}$ be coprime integers. As m and n are coprime, the primes p_i and q_j are distinct and so

$$\begin{aligned} f(mn) &= f\left(p_1^{k_1}\right)f\left(p_2^{k_2}\right)\cdots f\left(p_r^{k_r}\right)f\left(q_1^{l_1}\right)f\left(q_2^{l_2}\right)\cdots f\left(q_s^{l_s}\right) \\ &= f(m)f(n). \end{aligned}$$

Conversely, suppose that $f(mn) = f(m)f(n)$ whenever $\text{hcf}(m, n) = 1$. Then, for any integer $p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$ in prime decomposition,

The argument can be made watertight by the use of mathematical induction.

$$\begin{aligned} f\left(p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}\right) &= f\left(p_1^{k_1}\right) f\left(p_2^{k_2}\cdots p_r^{k_r}\right), \\ &\qquad \text{since } \text{hcf}\left(p_1^{k_1}, p_2^{k_2}\cdots p_r^{k_r}\right) = 1, \\ &= f\left(p_1^{k_1}\right) f\left(p_2^{k_2}\right) f\left(p_3^{k_3}\cdots p_r^{k_r}\right), \\ &\qquad \text{since } \text{hcf}\left(p_2^{k_2}, p_3^{k_3}\cdots p_r^{k_r}\right) = 1, \\ &= \cdots \\ &= f\left(p_1^{k_1}\right) f\left(p_2^{k_2}\right) f\left(p_3^{k_3}\right)\cdots f\left(p_r^{k_r}\right), \end{aligned}$$

and so f is multiplicative. ■

Exercise 1.4

If f is a multiplicative function prove that, with the exception of one function, $f(1) = 1$. What is the exceptional multiplicative function for which this does not hold?

The fact that σ and τ are multiplicative functions can be deduced from the following general result.

Proposition 1.5 *Generating new number-theoretic functions*

If f is a multiplicative function, then the number-theoretic function F defined by

$$F(n) = \sum_{d|n} f(d)$$

is also multiplicative.

Proof Let m and n be coprime integers. Then

$$F(mn) = \sum_{d|mn} f(d).$$

See Worked Exercise 4.13 of Book A, Chapter 1.

Now the factors of mn are the numbers of the form $d = rs$, where r and s are factors of m and n respectively. Therefore we can write

$$F(mn) = \sum_{r|m, s|n} f(rs),$$

Note that $\text{hcf}(r, s) = 1$ since any common factor of r and s would be a common factor of the coprime integers m and n.

and so, appealing to the multiplicative property of f, we have

$$F(mn) = \sum_{r|m, s|n} f(r)f(s).$$

On the other hand,

$$F(m)F(n) = \left(\sum_{r|m} f(r)\right)\left(\sum_{s|n} f(s)\right)$$

and, on multiplying out the product on the right, the sums for $F(mn)$ and $F(m)F(n)$ are seen to be the same. ■

The final step of the above proof, noting that the product of two sums is equal to the appropriate sum of individual products, may need amplification. To clarify this, we illustrate the equality of the summations by tracing the above proof for a particular case.

Take $m = 4$ and $n = 25$. We obtain $F(4 \times 25)$ by summing the values of f over the nine factors of 100:

$$\begin{aligned} F(100) = {} & f(1) + f(2) + f(4) + f(5) + f(10) \\ & + f(20) + f(25) + f(50) + f(100). \end{aligned}$$

Next we write each of the nine factors in its unique form as a factor of 4 multiplied by a factor of 25,

$$\begin{aligned} F(100) = {} & f(1 \times 1) + f(2 \times 1) + f(4 \times 1) + f(1 \times 5) + f(2 \times 5) \\ & + f(4 \times 5) + f(1 \times 25) + f(2 \times 25) + f(4 \times 25), \end{aligned}$$

and, appealing to the multiplicative property,

$$\begin{aligned} F(100) = {} & f(1)f(1) + f(2)f(1) + f(4)f(1) + f(1)f(5) + f(2)f(5) \\ & + f(4)f(5) + f(1)f(25) + f(2)f(25) + f(4)f(25). \end{aligned}$$

On the other hand,

$$F(4) = f(1) + f(2) + f(4) \quad \text{and} \quad F(25) = f(1) + f(5) + f(25),$$

and so

$$F(4)F(25) = (f(1) + f(2) + f(4))(f(1) + f(5) + f(25)).$$

When this is multiplied out we get a sum of nine terms that can be seen to be precisely the nine terms in the final expression for $F(100)$.

So far, the number of examples of multiplicative functions that we have met is very limited. There are two more functions that we can quickly add to our list. The constant function c and the identity function id, defined by

$$c(n) = 1 \quad \text{and} \quad \mathrm{id}(n) = n, \quad \text{for all } n \in \mathbb{Z},$$

are multiplicative because

$$c(mn) = 1 = c(m)c(n)$$

and

$$\mathrm{id}(mn) = mn = \mathrm{id}(m)\mathrm{id}(n).$$

Note that for each of these functions the multiplicative property holds for all integers m and n, regardless of whether or not they are coprime.

Now we can apply Proposition 1.5, with the role of f being played in turn by the functions c and id, in the knowledge that each will give rise to a multiplicative function:

$$\sum_{d|n} c(d) = \sum_{d|n} 1 = \tau(n)$$

and

$$\sum_{d|n} \mathrm{id}(d) = \sum_{d|n} d = \sigma(n).$$

So we have proved the following.

Corollary 1.6 *to Proposition 1.5*

The functions τ and σ are multiplicative.

Recalling that

$$\sigma(p^k) = (1 + p + p^2 + \cdots + p^k) = \frac{p^{k+1} - 1}{p - 1},$$

we can now use the multiplicative property to write down a formula for the value of $\sigma(n)$ in terms of the prime decomposition of n.

Proposition 1.7 *The prime decomposition formula for σ*

$$\begin{aligned}\sigma\left(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\right) &= \left(1 + p_1 + p_1^2 + \cdots + p_1^{k_1}\right) \\ &\quad \cdots \left(1 + p_r + p_r^2 + \cdots + p_r^{k_r}\right) \\ &= \prod_{1 \le i \le r} \frac{p_i^{k_i+1} - 1}{p_i - 1}\end{aligned}$$

For example,

$$\begin{aligned}\sigma\left(2^4 \times 5^2 \times 7\right) &= \frac{2^5 - 1}{2 - 1} \times \frac{5^3 - 1}{5 - 1} \times \frac{7^2 - 1}{7 - 1} \\ &= \frac{31}{1} \times \frac{124}{4} \times \frac{48}{6} \\ &= 31 \times 31 \times 8 = 7688.\end{aligned}$$

Exercise 1.5

(a) Show that the function $f : \mathbb{N} \to \mathbb{Z}$ given by

$$f(n) = \begin{cases} 0, & \text{if } n \text{ is even,} \\ n, & \text{if } n \text{ is odd,} \end{cases}$$

is multiplicative.

Describe the multiplicative function $F(n) = \sum_{d|n} f(d)$.

(b) Prove that for every natural number r, the function $\sigma_r : \mathbb{N} \to \mathbb{N}$ defined by

$$\begin{aligned} \sigma_r(n) &= \text{the sum of the } r\text{th powers of the positive factors of } n \\ &= \sum_{d|n} d^r, \end{aligned}$$

is multiplicative.

For example, $\sigma_4(6) = 1^4 + 2^4 + 3^4 + 6^4$.

Hint: apply Proposition 1.5 to a suitably chosen function.

Exercise 1.6

(a) An integer n is said to be **square-free** if it is not divisible by the square of any prime. Prove that if $n > 1$ is square-free then $\tau(n) = 2^r$ for some $r \geq 1$. Is the converse true, namely that if $\tau(n) = 2^r$ then n must be square-free?

(b) Find all integers n for which $\sigma(n) = 24$.

(c) Show that $\tau(n)$ is odd if, and only if, n is a square. For which integers n is $\sigma(n)$ odd?

(d) Show that $\prod_{d|n} d = n^{\tau(n)/2}$.

Hint: if d divides n then $\frac{n}{d}$ divides n and $d \times \frac{n}{d} = n$.

(e) Suppose that the natural number n has the property that $n + \sigma(n)$ is divisible by 3. Prove the following.

(i) If n is prime then $n \equiv 1 \pmod 6$.

(ii) The integer n cannot be the square of a prime.

(iii) If $n = pq$, where p and q are distinct odd primes with $p < q$, then $p = 3$ and $q \equiv 5 \pmod 6$.

(f) Prove that if f and g are multiplicative functions then the function defined for all natural numbers by $F(n) = f(n)g(n)$ is also multiplicative.

2 Perfect numbers

This section uses the σ function to look at numbers n that possess an interesting property: namely, that they equal the sum of their factors, excluding themselves.

2.1 Even perfect numbers

A number is said to be *perfect* if it is equal to the sum of its **parts**, meaning its positive factors excluding itself. The first two perfect numbers are $6 = 1 + 2 + 3$ and $28 = 1 + 2 + 4 + 7 + 14$. Perfect numbers can be defined in terms of the σ function as follows.

Definition 2.1 *Perfect numbers*

A natural number n is said to be **perfect** if it satisfies $\sigma(n) = 2n$.

Throughout history, mathematicians and philosophers have attached religious and mystical significance to the properties of numbers. Pride of place in this respect goes to the perfect numbers. The perfection of the creation is supposedly embodied in the facts that, according to the Bible, God took 6 days to create all things, and that the lunar cycle is 28 days long.

The search for perfect numbers has continued unabated since the time of the ancient Greeks. Only four were known to the Greeks, namely 6, 28, 496 and 8128. It is perhaps not surprising that they failed to extend this list, because the next two perfect numbers turn out to be 33 550 336 and 8 589 869 056. However, Euclid made a significant contribution in Book IX of his *Elements* when he revealed the following result, which we invite you to try your hand at proving.

Exercise 2.1

Prove that if k is a positive integer such that $2^k - 1$ is prime, then $n = 2^{k-1}\left(2^k - 1\right)$ is perfect.

Some 2000 years after Euclid had published his result, Euler proved that the perfect numbers characterised by Euclid are the only *even* ones.

Theorem 2.2 *Classification of even perfect numbers*

If k is a positive integer such that $2^k - 1$ is prime, then $n = 2^{k-1}(2^k - 1)$ is perfect. Furthermore, every even perfect number is of this form for some positive integer k.

Proof We have proved that such a number n is perfect and it remains to prove that, conversely, every even perfect number must be of this given form.

Let n be an even perfect number. Then $n = 2^r m$, where r is a positive integer and m is an odd positive integer. We can dismiss the case $m = 1$ because $\sigma(2^r) = 2^{r+1} - 1$ and this is not equal to 2×2^r. So we may assume that $m \geq 3$. Our goal is to show that $m = 2^{r+1} - 1$ and that it is prime.

Since 1 and m are distinct factors of m we can write $\sigma(m) = m + t$, where $t \geq 1$. Now,

$$\begin{aligned}\sigma(n) &= \sigma(2^r m) = \sigma(2^r)\sigma(m) = (2^{r+1} - 1)(m + t)\\ &= 2^{r+1}m - m + (2^{r+1} - 1)t\end{aligned}$$

and

$$2n = 2(2^r m) = 2^{r+1}m.$$

As n is perfect, $\sigma(n) = 2n$, and equating the above expressions for $\sigma(n)$ and $2n$ leads to

$$m = (2^{r+1} - 1)\, t.$$

Hence t is a factor of m and $t < m$ as 2^{r+1} is at least 4. Now if $t \neq 1$ then 1, t and m are three distinct factors of m, so that $\sigma(m) \geq m + t + 1$, which contradicts $\sigma(m) = m + t$. We are forced to conclude that $t = 1$, which gives $m = 2^{r+1} - 1$. Moreover $\sigma(m) = m + 1$, which confirms that m is prime. Therefore $n = 2^r\,(2^{r+1} - 1)$ and $2^{r+1} - 1$ is prime. This shows that n is of the correct form with $r = k - 1$ and completes the proof. ■

It may appear that Theorem 2.2 has solved the problem of even perfect numbers. We know exactly where to find them all – or do we?

2.2 Mersenne primes

Theorem 2.2 leaves one burning question. For which integers k is $2^k - 1$ prime? Each prime of this form leads to a perfect number and all even perfect numbers are obtained in this way. The list of candidates is quickly whittled down when we observe the identity

$$2^{rs} - 1 = (2^r - 1)\left(2^{r(s-1)} + 2^{r(s-2)} + 2^{r(s-3)} + \cdots + 2^{2r} + 2^r + 1\right),$$

which shows that $2^k - 1$ is composite whenever $k = rs$ is composite. The question therefore becomes, for which primes p is $2^p - 1$ prime? For the first four primes, the value of $2^p - 1$ turns out to be prime, giving the first four perfect numbers, as shown in Table 2.1.

Table 2.1 Four perfect numbers

p	$2^p - 1$	$2^{p-1}(2^p - 1)$
2	3	$2 \times 3 = 6$
3	7	$4 \times 7 = 28$
5	31	$16 \times 31 = 496$
7	127	$64 \times 127 = 8128$

Early writers believed that $2^p - 1$ is prime for all primes p, but this was eventually contradicted in 1536 when the factorisation for the fifth number in this sequence was discovered:

$$2^{11} - 1 = 2047 = 23 \times 89.$$

By 1588 Cataldi (1548–1626), making use of a table of primes and testing for division by each prime less than the square root, had confirmed that $2^{13} - 1$, $2^{17} - 1$ and $2^{19} - 1$ are primes. He went on to claim that $2^p - 1$ is prime for the next four prime values of p, namely 23, 29, 31 and 37. However Fermat (1601?–1665) proved, around 1640, that two of these assertions were false by establishing that 47 divides $2^{23} - 1$ and that 223 divides $2^{37} - 1$.

You can learn more about Mersenne in the *History Reader* for this topic.

In 1644, Marin Mersenne (1588–1648) claimed that the only primes $p \leq 257$ for which $2^p - 1$ is also prime are 2, 3, 5, 7, 13, 17, 19, 31, 67, 127 and 257. Though this list was later found to contain some errors, numbers of the form $2^p - 1$ became known as the *Mersenne numbers*, M_p.

Definition 2.3 *Mersenne numbers*

The numbers $M_p = 2^p - 1$, where p is prime, are known as **Mersenne numbers**. When M_p is prime it is referred to as a **Mersenne prime**.

In 1732 Euler confirmed that M_{29} is composite and 40 years later he established that M_{31} is indeed prime by eliminating all possible factors. For some 200 years M_{19} had stood as the largest known prime number but now a much larger one, M_{31}, was to hand.

To keep track of how the even perfect numbers are coming along, M_{19} and M_{31} provide us with the seventh and eighth perfect numbers:

$$2^{18}\left(2^{19}-1\right) = 137\,438\,691\,328,$$
$$2^{30}\left(2^{31}-1\right) = 2\,305\,843\,008\,139\,952\,128.$$

By 1811 no larger prime number had been found and a distinguished mathematician of that time, Barlow (1776–1862), asserted pessimistically in his book *Theory of Numbers* that 'M_{31} is the greatest prime that will ever be found'. His reasoning was that the search for larger primes had curiosity value rather than real use, and the sheer size of M_{31} would deter a search for anything larger. How wrong he was. A mere 65 years later, the French mathematician Edouard Lucas (1842–1891) proved that

$$M_{127} = 170\,141\,183\,460\,469\,231\,731\,687\,303\,715\,884\,105\,727$$

is prime, and with it a ninth perfect number, $2^{126}\left(2^{127}-1\right)$, which has 77 digits, had been found. Lucas also uncovered the first mistake in Mersenne's list when he proved that M_{67} is in fact composite. His proof demonstrated the lack of primality without producing the actual factors. That had to wait until 1903 when, at a meeting of the American Mathematical Society, a professor called F.N. Cole delivered a talk entitled 'On the factorisation of large numbers'. His 'talk' consisted of calculating $2^{67}-1$ by hand on one half of the blackboard and then, on the other half, working out the product

$$193\,707\,721 \times 761\,838\,257\,287.$$

The two agreed. He sat down to prolonged applause, having said not a single word!

By 1914 three omissions in Mersenne's list had been found: M_{61}, M_{89} and M_{107} had all been shown to be prime. But M_{127} remained the largest known Mersenne prime until 1952, when an American, R. Robinson, using what was then regarded as a large computer, showed that M_{521}, M_{607}, M_{1279}, M_{2203} and M_{2281} are all prime. In 1996, a collaborative project entitled GIMPS (Great Internet Mersenne Prime Search) was set up to enable computer owners to contribute spare processing power to search for Mersenne primes. By 2014 this effort meant that the list had reached 48 Mersenne primes, and therefore 48 perfect numbers. These are given in Table 2.2.

Though M_{127} was the ninth Mersenne prime discovered, it is the twelfth in ascending order.

Table 2.2 Mersenne primes

Prime p	Number of digits in M_p	Perfect number	Date discovered
2	1	$2(2^2-1)$	Known to Euclid
3	1	$2^2(2^3-1)$	Known to Euclid
5	2	$2^4(2^5-1)$	Known to Euclid
7	3	$2^6(2^7-1)$	Known to Euclid
13	4	$2^{12}(2^{13}-1)$	1456
17	6	$2^{16}(2^{17}-1)$	1588
19	6	$2^{18}(2^{19}-1)$	1588
31	10	$2^{30}(2^{31}-1)$	1772
61	19	$2^{60}(2^{61}-1)$	1883
89	27	$2^{88}(2^{89}-1)$	1911
107	33	$2^{106}(2^{107}-1)$	1914
127	39	$2^{126}(2^{127}-1)$	1876
521	157	$2^{520}(2^{521}-1)$	1952
607	183	$2^{606}(2^{607}-1)$	1952
1279	386	$2^{1278}(2^{1279}-1)$	1952
2203	664	$2^{2202}(2^{2203}-1)$	1952
2281	687	$2^{2280}(2^{2281}-1)$	1952
3217	969	$2^{3216}(2^{3217}-1)$	1957
4253	1281	$2^{4252}(2^{4253}-1)$	1961
4423	1332	$2^{4422}(2^{4423}-1)$	1961
9689	2917	$2^{9688}(2^{9689}-1)$	1963
9941	2993	$2^{9940}(2^{9941}-1)$	1963
11 213	3376	$2^{11\,212}(2^{11\,213}-1)$	1963
19 937	6002	$2^{19\,936}(2^{19\,937}-1)$	1971
21 701	6533	$2^{21\,700}(2^{21\,701}-1)$	1978
23 209	6987	$2^{23\,208}(2^{23\,209}-1)$	1979
44 497	13 395	$2^{44\,496}(2^{44\,497}-1)$	1979
86 243	25 962	$2^{86\,242}(2^{86\,243}-1)$	1982
110 503	33 265	$2^{110\,502}(2^{110\,503}-1)$	1988
132 049	39 751	$2^{132\,048}(2^{132\,049}-1)$	1983
216 091	65 050	$2^{216\,090}(2^{216\,091}-1)$	1985
756 839	227 832	$2^{756\,838}(2^{756\,839}-1)$	1992
859 433	258 716	$2^{859\,432}(2^{859\,433}-1)$	1994
1 257 787	378 632	$2^{1\,257\,786}(2^{1\,257\,787}-1)$	1996
1 398 269	420 921	$2^{1\,398\,268}(2^{1\,398\,269}-1)$	1996
2 976 221	895 932	$2^{2\,976\,220}(2^{2\,976\,221}-1)$	1997
3 021 377	909 526	$2^{3\,021\,376}(2^{3\,021\,377}-1)$	1998
6 972 593	2 098 960	$2^{6\,972\,592}(2^{6\,972\,593}-1)$	1999
13 466 917	4 053 946	$2^{13\,466\,916}(2^{13\,466\,917}-1)$	2001
20 996 011	6 320 430	$2^{20\,996\,010}(2^{20\,996\,011}-1)$	2003
24 036 583	7 235 733	$2^{24\,036\,582}(2^{24\,036\,583}-1)$	2004
25 964 951	7 816 230	$2^{25\,964\,950}(2^{25\,964\,951}-1)$	2005
30 402 457	9 152 052	$2^{30\,402\,456}(2^{30\,402\,457}-1)$	2005
32 582 657	9 808 358	$2^{32\,582\,656}(2^{32\,582\,657}-1)$	2006
37 156 667	11 185 272	$2^{37\,156\,666}(2^{37\,156\,667}-1)$	2008
42 643 801	12 837 064	$2^{42\,643\,800}(2^{42\,643\,801}-1)$	2009
43 112 609	12 978 189	$2^{43\,112\,608}(2^{43\,112\,609}-1)$	2008
57 885 161	17 425 170	$2^{57\,885\,160}(2^{57\,885\,161}-1)$	2013

2.3 Some results

From the unfolding story you might not be surprised to learn that we do not know yet whether or not there exist infinitely many Mersenne primes, and hence infinitely many even perfect numbers. Curiously, we are no further forward in the opposite direction either. It remains unknown whether or not infinitely many of the Mersenne numbers M_p are composite. One hope for the infinitude of the Mersenne primes arose from the observation that for the first four such numbers, namely 3, 7, 31 and 127, the numbers M_3, M_7, M_{31}, and M_{127} had also turned out to be prime. Could it be generally true that if M_p is prime then M_{M_p} is prime? If so, we would immediately have infinitely many Mersenne primes. But this hope was dashed when, in 1953, a computer confirmed that

$$M_{M_{13}} = 2^{M_{13}} - 1 = 2^{8191} - 1$$

is composite.

As the challenge to determine the primality or otherwise of M_p has persisted, a variety of methods and results for attacking particular values of p have emerged. As illustrations we introduce a couple of the simpler ones here.

Exercise 2.2

Suppose that p and $q = 2p + 1$ are both prime. By applying FLT to 2^{q-1}, show that either q divides M_p or q divides $M_p + 2$.

The result of Exercise 2.2 can be followed through to the conclusion that q divides M_p when $p \equiv 3 \pmod 4$. We record this result but postpone the proof of the missing final step until Chapter 10, when it can be deduced quickly from results established there.

Proposition 2.4

If p and $2p + 1$ are primes, where $p \equiv 3 \pmod 4$, then $2p + 1$ divides M_p.

To illustrate the use of Proposition 2.4, note that $11 \equiv 3 \pmod 4$ and $2 \times 11 + 1 = 23$ is prime. We conclude that 23 is a factor of M_{11}, a fact which we already knew. But in just the same way, Proposition 2.4, together with a brief glance at a table of primes, tells us that 47 is a factor of M_{23}, 167 is a factor of M_{83}, 263 is a factor of M_{131}, 359 is a factor of M_{179}, 479 is a factor of M_{239} and 503 is a factor of M_{251}. So this one result has confirmed seven of the composite values in Mersenne's list.

The second result, which was known to Fermat, is not quite so dramatic but it severely restricts the potential factors of a given M_p.

Proposition 2.5

Any prime factor of M_p, where p is an odd prime, is of the form $2kp + 1$ for some positive integer k.

Proof Let q be a prime factor of $M_p = 2^p - 1$. Then $2^p \equiv 1 \pmod{q}$. But we also have, from FLT, that $2^{q-1} \equiv 1 \pmod{q}$. Now let the order of 2 modulo q be c, so that both p and $q - 1$ are multiples of c. But as p is a prime and $c > 1$ (since $2^1 \equiv 1 \pmod{q}$ is not possible), it follows that $c = p$. Hence $q - 1$ is a multiple of p.

See Proposition 2.4 of Book A, Chapter 4.

Finally, writing $q - 1 = tp$, for some integer t, gives $q = tp + 1$. But if the integer t is odd then the prime q would be even, which certainly cannot be the case. Hence t is a positive even integer, and putting $t = 2k$ gives $q = 2kp + 1$, as claimed. ■

To illustrate how Proposition 2.5 might be applied, consider M_{19}. The prime factors of M_{19} must be of the form $38k + 1$, and if M_{19} is not itself prime then it must have a factor that does not exceed the square root of M_{19} (this is approximately $2^{19/2}$, which is close to 724). There are nineteen numbers of the form $38k + 1$, with k a natural number, that do not exceed 724, namely 39, 77, 115, 153, ..., 723, and reference to a table of primes reveals that just six of these are prime, namely 191, 229, 419, 457, 571 and 647.

Hence, to establish the primality of M_{19}, it is sufficient to show that it does not have one of these six primes as a factor.

Exercise 2.3

Applying Proposition 2.5 to test the primality of M_{17}, it can be shown that, to be composite, M_{17} must be divisible by one of four primes. Find these four primes.

We have concentrated on the even perfect numbers with good reason. No odd perfect number is known; it is generally believed that odd numbers cannot be perfect, but nobody has managed to prove this. The search for odd perfect numbers has, however, not been without interest. Euler made the first real start when he proved the following result concerning the prime decomposition that an odd perfect number must have.

Proposition 2.6 *Euler's form for an odd perfect number*

If n is an odd perfect number then

$$n = p^k p_1^{2k_1} p_2^{2k_2} \cdots p_r^{2k_r},$$

where the p_is are distinct primes and $p \equiv k \equiv 1 \pmod{4}$.

The exceptional prime p has been listed first but is not necessarily the smallest prime in the decomposition. Proof of Proposition 2.6 is beyond the scope of this module.

This result says that with one exception all the exponents in the prime decomposition of n must be even, and the exceptional prime and its exponent are both of the form $4t+1$ (though not necessarily for the same t). This result has, nowadays, been developed much further. For example, it is now known that the total number of distinct primes involved must be at least nine, one of the prime factors must be greater than 10^8 and the number n itself must have more than 300 digits.

We cannot leave this subject matter without saying a few words about the *Fermat numbers*. Along with the Mersenne numbers, which have the form 2^n-1, it seems natural to consider numbers of the form 2^n+1. Can such numbers be prime? Fermat was aware that if the exponent n has an odd factor (other than 1) then 2^n+1 must be composite, as can be seen from the identity

$$2^{(2r+1)s}+1=(2^s+1)\Big(2^{2rs}-2^{(2r-1)s}+2^{(2r-2)s}-\cdots-2^s+1\Big).$$

To have a chance of being prime the exponent n must therefore be a power of 2, and so Fermat turned his attention to numbers of the form $2^{2^n}+1$.

Definition 2.7 *Fermat numbers*

The numbers $F_n=2^{2^n}+1$, where n is a non-negative integer, are known as **Fermat numbers**.

2^{2^n} means $2^{(2^n)}$, not $\left(2^2\right)^n$.

The first five Fermat numbers,

$$F_0=3,\quad F_1=5,\quad F_2=17,\quad F_3=257\quad\text{and}\quad F_4=65\,537,$$

are primes.

In a letter to Mersenne, Fermat expressed his belief that all the numbers F_n are prime, although he admitted that he could not provide a proof. It was Euler who provided the counter-example: that 641 divides F_5. By 2010, all the numbers F_n for n from 5 to 32 had been shown to be composite. The prime decomposition

$$F_6=274\,177\times 67\,280\,421\,310\,721$$

was known in 1880, but it required a computer to uncover, in 1970, the next one:

$$F_7=59\,649\,589\,127\,497\,217\times 5\,704\,689\,200\,685\,129\,054\,721.$$

The prime decomposition of F_8 was found in 1980, followed by that of F_{11} in 1989. F_{11} has two relatively small prime factors:

$$F_{11}=319\,489\times 974\,849\times p_1\times p_2\times p_3,$$

where the primes p_1, p_2 and p_3 have 21, 22 and 564 digits respectively. The prime decomposition of F_9 was found in 1990, and that of F_{10} in 1995. As with the search for Mersenne primes, there is now a distributed computing project called Fermat Search, dedicated to the task of finding factors of Fermat numbers. Even so, by 2010 only F_0 to F_{11} had been completely factorised. However, at least one prime factor of F_n is known for all

$n \leq 32$, except for $n = 20$ and $n = 24$. Although no actual prime factor had been found in either case, F_{20} and F_{24} were shown to be composite. At the time of writing, that leaves F_{33} as the next Fermat number that may yet turn out to be prime.

Many more Fermat numbers have been shown to be composite by various methods, which often hinge on exhibiting some prime that must be a factor. The largest Fermat number known to be composite at the time of writing is $F_{2747497}$, which was shown, in 2013, to be divisible by $57 \times 2^{2747499} + 1$. Compared with the size of $F_{2747497}$ itself, this is a very, very small factor, and yet it is a number with over 800 000 digits!

Exercise 2.4

(a) Prove that n is perfect if, and only if,

$$\sum_{d|n} \frac{1}{d} = 2.$$

Deduce that if n is perfect and d is a factor of n other than n itself, then d is not perfect.

These properties hold for all perfect numbers where, in addition, k and $2^k - 1$ are odd primes.

(b) Let $n = 2^{k-1}\left(2^k - 1\right)$, where the integer $k \geq 3$ is odd. Show that:

(i) $n \equiv 4 \pmod 6$

(ii) $n \equiv 1 \pmod 9$

(iii) n is a triangular number.

(c) An integer $n \geq 1$ is said to be **abundant** if $\sigma(n) > 2n$ and **deficient** if $\sigma(n) < 2n$.

(i) Classify each of the numbers from 24 to 30 inclusive as abundant, deficient or perfect.

(ii) Show that if p is prime and $r \geq 1$, then p^r is deficient.

(iii) Show that, with one exception, any product of two distinct primes is deficient.

(iv) Find an odd abundant number. This is not easy!

Hint: you will have to involve three different odd primes.

(v) Show that if m is abundant and $\text{hcf}(m, n) = 1$, then mn is abundant.

(vi) Show that for integers $k > 1$, $n = 2^{k-1}\left(2^k - 1\right)$ is abundant whenever $2^k - 1$ is not prime.

(d) A pair of natural numbers m and n that satisfies

$$\sigma(m) = \sigma(n) = m + n$$

is said to be an **amicable pair**.

(i) Verify that the following pairs are amicable:

220 and 284 (Pythagoras, 500 BC),
$17\,296 = 2^4 \times 23 \times 47$ and $18\,416 = 2^4 \times 1151$ (Fermat, 1636),
$2^7 \times 191 \times 383$ and $2^7 \times 73\,727$ (Descartes, 1638).

Note that 1151, 191, 383 and 73 727 are primes.

(ii) Show that a prime p cannot be one of an amicable pair.

3 Euler's ϕ-function

This section uses another multiplicative function to revisit the idea of order from Book A, Chapter 4.

3.1 Euler's Theorem

Let us look again at the *order* of an integer. In Section 2 of Book A, Chapter 4 we had the following definition.

If p is prime and $\text{hcf}(a, p) = 1$, then the **order** of a modulo p is the least positive integer c such that $a^c \equiv 1 \pmod{p}$.

In this section we are going to investigate the notion of order for composite moduli. In Table 3.1 we have calculated successive powers of integers a, modulo 10, in an attempt to find the order of a by discovering a power that is congruent to 1.

Table 3.1 Powers modulo 10

a	1	2	3	4	5	6	7	8	9
a^2	1	4	9	6	5	6	9	4	1
a^3	1	8	7	4	5	6	3	2	9
a^4	1	6	1	6	5	6	1	6	1
a^5	1	2	3	4	5	6	7	8	9

We stopped at the fifth power in the realisation that the value 1 is not going to arise for $a = 2$, 4, 5, 6 and 8, where the powers are already cycling around values other than 1. What the table tells us is that the numbers 1, 3, 7 and 9 have respective orders 1, 4, 4 and 2 modulo 10, but the concept of order does not apply to the remaining integers. We note the four that have an order are precisely the four that are coprime to the modulus 10.

Let us turn now to a general modulus n and suppose that the number a has an order modulo n, that is, there is an integer $r \geq 1$ such that

$a^r \equiv 1 \pmod{n}$. This means that $a^r - 1 = kn$, for some integer k. Now $\text{hcf}(a, n)$ divides a, and therefore a^r, and it divides n, and so divides $a^r - kn = 1$. Thus $\text{hcf}(a, n) = 1$, and so a is coprime to n.

All this suggests that we should be concentrating on the set of least positive residues of n that are coprime to n. Before defining this set it will be useful to define the following number-theoretic function, which gives the size of this set.

Definition 3.1 *Euler's ϕ-function*

For each integer $n \geq 1$, define $\phi(n)$ to be the number of natural numbers not exceeding n that are coprime to n.

Note that $\phi(1) = 1$ since $\text{hcf}(1, 1) = 1$.

For example, the integers a in the range $1 \leq a \leq 10$ for which $\text{hcf}(a, 10) = 1$ are 1, 3, 7 and 9, four in number, and so $\phi(10) = 4$. In terms of residue classes, $\phi(n)$ is the number of classes that consist of integers coprime to n. Selection of one member from each such class produces a *reduced set of residues*.

Definition 3.2 *Reduced set of residues*

A **reduced set of residues modulo** n is a set of integers $\{a_1, a_2, a_3, \ldots, a_{\phi(n)}\}$ each of which is coprime to n and no two of which are congruent modulo n.

The reduced set of residues in which each is a positive integer less than or equal to n is called the **reduced set of least positive residues modulo** n.

As $\text{hcf}(0, n) = n$, 0 does not appear in a reduced set of residues modulo n for any $n > 1$.

Exercise 3.1

Complete the following table giving, for $n = 1$ to 10, the reduced set of least positive residues modulo n and the value of $\phi(n)$.

n	Reduced set of residues	$\phi(n)$
1		
2		
3		
4		
5		
6		
7		
8		
9		
10	{1,3,7,9}	4

To determine $\phi(n)$ for $n > 1$, we need to know how many numbers in the set $\{1, 2, 3, \ldots, n-1\}$ are coprime to n. If n happens to be prime they are all coprime to n, and so $\phi(n) = n - 1$. On the other hand, if n is composite then one of the listed numbers is a factor $d > 1$ of n, and as $\text{hcf}(d, n) = d$, it is not coprime to n. In this case $\phi(n) \leq n - 2$. The conclusion from these remarks is that n is prime if, and only if, $\phi(n) = n - 1$.

Since $n > 1$ we may exclude n from consideration since $\text{hcf}(n, n) = n \neq 1$.

We will return to investigate properties of the ϕ-function shortly, but in the meantime let us head straight for our main result of this section: Euler's generalisation of FLT. For a composite modulus n, and with attention restricted to integers a coprime to n, we are interested in finding exponents c such that $a^c \equiv 1 \pmod{n}$. The smallest natural number c satisfying this congruence will be the order of a modulo n. To garner more numerical evidence, we have calculated various powers of a reduced set of residues modulo 9 and the results are shown in Table 3.2.

Table 3.2 Powers modulo 9

a	1	2	4	5	7	8
a^2	1	4	7	7	4	1
a^3	1	8	1	8	1	8
a^4	1	7	4	4	7	1
a^5	1	5	7	2	4	8
a^6	1	1	1	1	1	1

What should first catch the eye about this table is the row of sixth powers: $a^6 \equiv 1 \pmod{9}$ for each integer a coprime to 9. A shrewd observer will also have noted that the exponent 6 just happens to be $\phi(9)$. There is nothing special about the number 9; what is illustrated here is true more generally.

Theorem 3.3 *Euler's Theorem – a generalisation of FLT*

If n is a natural number and a is any integer with $\text{hcf}(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod{n}$.

Notice that when n is a prime p, then, since $\phi(p) = p - 1$, this result says that $a^{p-1} \equiv 1 \pmod{p}$, which is precisely FLT. So this result is indeed a generalisation.

The proof that follows ought to look familiar, for it is little more than a re-run of our proof of FLT. However, where the latter worked with a set of non-zero residues modulo p, we now base the construction on a reduced set of residues modulo n.

Proof Let $\{b_1, b_2, b_3, \ldots, b_{\phi(n)}\}$ be a reduced set of residues modulo n. Consider the set obtained by multiplying each by a:

$$\{ab_1, ab_2, ab_3, \ldots, ab_{\phi(n)}\}.$$

We claim that this is also a reduced set of residues modulo n. Certainly no two of the $\phi(n)$ integers in this set are congruent modulo n because from $ab_i \equiv ab_j \pmod{n}$ and $\text{hcf}(a, n) = 1$, cancellation would give $b_i \equiv b_j \pmod{n}$, which can occur only when $i = j$. Moreover, each integer of the set is coprime to n because, if some prime p divides both ab_i and n then, appealing to Euclid's Lemma, p divides either a or b_i. But p cannot divide both a and n for that would contradict $\text{hcf}(a, n) = 1$, and similarly p cannot divide both b_i and n because that contradicts $\text{hcf}(b_i, n) = 1$. It follows that such a prime p cannot exist and therefore $\text{hcf}(ab_i, n) = 1$.

Each b_i is congruent to some ab_j modulo n.

As each of the two sets is a reduced set of residues modulo n their products are congruent modulo n, that is,

$$ab_1 ab_2 ab_3 \cdots ab_{\phi(n)} \equiv b_1 b_2 b_3 \cdots b_{\phi(n)} \pmod{n}.$$

Now since $\text{hcf}(b_i, n) = 1$, we can cancel each b_i in turn to obtain $a^{\phi(n)} \equiv 1 \pmod{n}$, as claimed. ■

To illustrate the above proof, let us trace it through for the case $n = 15$ and $a = 7$. We obtain a reduced set of residues modulo 15 by taking the set of least positive residues of 15 and eliminating those that are multiples of either 3 or 5, leaving the set of $\phi(15) = 8$ integers, $\{1, 2, 4, 7, 8, 11, 13, 14\}$. The proof showed that we obtain the same set of eight integers, modulo 15, when we multiply each by 7. In fact

$$7 \times 1 \equiv 7, \quad 7 \times 2 \equiv 14, \quad 7 \times 4 \equiv 13, \quad 7 \times 7 \equiv 4,$$
$$7 \times 8 \equiv 11, \quad 7 \times 11 \equiv 2, \quad 7 \times 13 \equiv 1, \quad 7 \times 14 \equiv 8.$$

Therefore

$$7^8 \times (1 \times 2 \times 4 \times 7 \times 8 \times 11 \times 13 \times 14) \equiv$$
$$1 \times 2 \times 4 \times 7 \times 8 \times 11 \times 13 \times 14 \pmod{15},$$

and after cancellation by the least positive residues, which is permissible since each of the eight numbers is coprime to 15, we obtain $7^8 \equiv 1 \pmod{15}$.

Exercise 3.2

Determine $\phi(14)$ and confirm that

$$3^{\phi(14)} \equiv 1 \pmod{14} \quad \text{and} \quad 5^{\phi(14)} \equiv 1 \pmod{14}.$$

3.2 The order of an integer to a composite modulus

With the benefit of Euler's Theorem, we are now in a position to give a general definition of the *order* of a number.

Definition 3.4 *The order of an integer modulo n*

If $n \geq 1$ and $\text{hcf}(a, n) = 1$, then the **order** of a modulo n is the least positive integer c such that $a^c \equiv 1 \pmod{n}$.

We were already aware that the condition $\text{hcf}(a, n) = 1$ is necessary for the existence of an integer c with $a^c \equiv 1 \pmod{n}$. Euler's Theorem shows that the condition is also sufficient, since $a^{\phi(n)} \equiv 1 \pmod{n}$. So $c \leq \phi(n)$.

Since $a^k \equiv b^k \pmod{n}$ whenever $a \equiv b \pmod{n}$, it can be seen that congruent integers have the same order. Hence to find all orders modulo n it suffices to find the orders of the $\phi(n)$ integers in a reduced set of residues modulo n.

Exercise 3.3

Find all the possible orders of integers modulo 14.

In Book A, Chapter 4, we established that if $a^k \equiv 1 \pmod{p}$, for some prime p, then k must be a multiple of the order of a. It should come as no surprise to learn that this result also generalises.

Proposition 2.4 of Book A, Chapter 4.

Proposition 3.5

If the integer a has order c modulo n then $a^k \equiv 1 \pmod{n}$ if, and only if, k is a multiple of c. In particular, the order c divides $\phi(n)$.

Proof If k is a multiple of c, say $k = cr$, then

$$a^k = a^{cr} = (a^c)^r \equiv 1^r \equiv 1 \pmod{n}.$$

On the other hand, suppose that $a^k \equiv 1 \pmod{n}$. Dividing k by c, the Division Algorithm gives integers q and r such that $k = qc + r$, where $0 \leq r < c$. Then

$$1 \equiv a^k \equiv a^{qc+r} \equiv (a^c)^q a^r \equiv (1)^q a^r \equiv a^r \pmod{n}.$$

This forces $r = 0$, for otherwise $a^r \equiv 1 \pmod{n}$ contradicts c being the least positive integer with the property $a^c \equiv 1 \pmod{n}$. So $k = qc$, and k is a multiple of c, as claimed.

As $a^{\phi(n)} \equiv 1 \pmod{n}$, it follows that $\phi(n)$ is a multiple of the order c. ■

Proposition 3.5 simplifies the computational task of determining the order of an integer modulo n, because it tells us that we can confine attention to those exponents that are factors of $\phi(n)$.

Worked Exercise 3.6

Determine the order of 5 modulo 38.

Solution

To find $\phi(38)$ it suffices to count how many integers in the set $\{1, 2, 3, \ldots, 36, 37\}$ are coprime to 38. As 38 has prime decomposition 2×19, the integers coprime to 38 are precisely those that are not multiples of either 2 or 19. This leaves the odd numbers with the exception of 19, of which there are eighteen. So $\phi(38) = 18$. The order of 5 modulo 38 is therefore a factor of 18 and consequently is one of 1, 2, 3, 6, 9 and 18. Working modulo 38, we have

$$5^1 = 5,$$
$$5^2 = 25,$$
$$5^3 = 125 \equiv 11,$$
$$5^6 = 5^3 \times 5^3 \equiv 11 \times 11 \equiv 7,$$
$$5^9 = 5^6 \times 5^3 \equiv 7 \times 11 \equiv 1.$$

Hence the order of 5 modulo 38 is 9.

Exercise 3.4

Determine:

(a) the order of 4 modulo 21

(b) the order of 7 modulo 22.

3.3 Applications of Euler's Theorem

We have seen numerous applications of FLT that assist in the determination of remainders when large exponents are involved. Euler's Theorem can be used in the same way, but with the extra dimension that we need not be working to a prime modulus.

Worked Exercise 3.7

Determine the remainder when 23^{23} is divided by 15.

Solution

Since $\text{hcf}(23, 15) = 1$ and $\phi(15) = 8$, Euler's Theorem tells us that $23^8 \equiv 1 \pmod{15}$.

Therefore

$$\begin{aligned} 23^{23} = (23^8)^2 \times 23^7 \equiv 23^7 \equiv 8^7 = 64^3 \times 8 \\ \equiv 4^3 \times 8 = 16 \times 4 \times 8 \equiv 1 \times 32 \equiv 2 \pmod{15}, \end{aligned}$$

and the remainder on dividing 23^{23} by 15 is 2.

Exercise 3.5

Confirm that $\phi(100) = 40$ and hence determine the final two digits of the following numbers.

(a) 17^{83}

(b) 19^{39}

Hint: perhaps consider the congruence $19x \equiv 1 \pmod{100}$.

The sequence of integers

$$1,\ 11,\ 111,\ 1\,111,\ 11\,111, \ldots,$$

the nth term of which is the number comprising n 1s, has aroused a great deal of interest. The numbers in this sequence are called **repunits** (pronounced 'rep units') and, just as with the Mersenne numbers, much effort has been expended on attempts to factorise them. If we denote the repunit with n 1s by R_n then it turns out that if n is composite, so too is R_n. In fact, if $n = rs$ then each of R_r and R_s is a factor of R_{rs}, as can be seen from the result of Exercise 3.6(a) below.

For example,

$$\begin{aligned} R_6 &= 111\,111 \\ &= 11 \times 10\,101 = R_2 \times 10\,101 \\ &= 111 \times 1001 = R_3 \times 1001. \end{aligned}$$

Prime repunits are rare. Table 3.3 shows the prime decompositions of R_p for primes p up to 30, and you will notice that just three of these turn out to be prime.

Table 3.3 Prime decompositions of small repunits

p	R_p
2	11
3	3×37
5	41×271
7	239×4649
11	$21\,649 \times 513\,239$
13	$53 \times 79 \times 265\,371\,653$
17	$2\,071\,723 \times 5\,363\,222\,357$
19	$1\,111\,111\,111\,111\,111\,111$
23	$11\,111\,111\,111\,111\,111\,111\,111$
29	$3191 \times 16\,763 \times 43\,037 \times 62\,003 \times 77\,843\,839\,397$

Before progressing, let us record two simple facts about repunits that will prove useful in what follows. First, suppose we multiply the repunit R_n by 9:

$$9R_n = \overbrace{999\ldots99}^{n\text{ 9s}} = 10^n - 1.$$

This gives us a new expression for R_n, shown below.

Repunit property 1

$$R_n = \frac{10^n - 1}{9}$$

Second, consider the repunit R_{m+n}.

$$\begin{aligned} R_{m+n} &= \overbrace{111\ldots1}^{m\text{ 1s}}\overbrace{1\ldots11}^{n\text{ 1s}} \\ &= \overbrace{111\ldots1}^{m\text{ 1s}}\overbrace{0\ldots00}^{n\text{ 0s}} + \overbrace{1\ldots11}^{n\text{ 1s}} \\ &= R_m \times 10^n + R_n \end{aligned}$$

Repunit property 2

$$R_{m+n} = R_m \times 10^n + R_n$$

If you were not convinced by the argument leading to property 2, you can prove it in a straightforward fashion by substituting the expressions for R_{m+n}, R_m and R_n given by property 1.

One curiosity associated with repunits is explained in the following worked exercise: every odd integer that is not divisible by 5 has a multiple that is a repunit.

Worked Exercise 3.8

Show that if the integer n is coprime to 10 then n has a multiple each of whose digits is a 1.

Solution

Suppose n is such that $\text{hcf}(n, 10) = 1$. Since $\text{hcf}(9, 10) = 1$ we have that $\text{hcf}(9n, 10) = 1$ and so we can apply Euler's Theorem:

$$10^{\phi(9n)} \equiv 1 \,(\text{mod } 9n),$$

that is,

$$10^{\phi(9n)} - 1 = k \times 9n, \quad \text{for some integer } k.$$

Dividing both sides of this equation by 9, we obtain

$$R_{\phi(9n)} = kn,$$

which confirms the claim that n has a multiple that is a repunit, namely $R_{\phi(9n)}$.

Exercise 3.6

(a) Use mathematical induction to prove that R_{mn} is divisible by R_n, for all integers $m \geq 1$.

(b) Let $p > 5$ be prime.

(i) Show that p divides R_{p-1}.

Hint: consider the number $10^{p-1} - 1$.

(ii) Use the Division Algorithm to show that if R_n is the smallest repunit that is a multiple of p, then n divides $p - 1$.

(iii) Find the smallest repunit that is a multiple of (1) 7 and (2) 13.

We will go through one final example using Euler's Theorem.

Worked Exercise 3.9

Show that if $\text{hcf}(n, 42) = 1$ then $n^6 \equiv 1 \pmod{7 \times 8 \times 9}$.

Solution

Using what should be a familiar approach, we aim to show that n^6 is congruent to 1 for each of the pairwise coprime moduli 7, 8 and 9. Since $\phi(7) = 6$ and $\phi(9) = 6$, Euler's Theorem gives

$$n^6 \equiv 1 \pmod{7}, \text{ whenever } \text{hcf}(n, 7) = 1, \text{ and}$$
$$n^6 \equiv 1 \pmod{9}, \text{ whenever } \text{hcf}(n, 9) = 1, \text{ that is,}$$
$$\text{whenever } \text{hcf}(n, 3) = 1.$$

Finally, since the square of any odd integer is congruent to 1 modulo 8 we have

$$n^6 \equiv 1 \pmod{8}, \text{ for all integers } n \text{ such that } \text{hcf}(n, 2) = 1.$$

Now the given condition $\text{hcf}(n, 42) = 1$ amounts to $\text{hcf}(n, 2) = 1$, $\text{hcf}(n, 3) = 1$ and $\text{hcf}(n, 7) = 1$, and so all three congruences hold and $n^6 \equiv 1 \pmod{7 \times 8 \times 9}$ follows.

Any odd integer is of the form $2n + 1$ and

$$(2n+1)^2 = 4n^2 + 4n + 1 = 4n(n+1) + 1.$$

As either n or $n + 1$ is even, the square is congruent to 1 modulo 8.

Exercise 3.7

Show that $n^{33} \equiv n \pmod{15 \times 16 \times 17}$ for every odd integer n.

Exercise 3.8

(a) (i) Write down a reduced set of residues modulo 20.

(ii) Show that $\{5, 5^2, 5^3, 5^4, 5^5, 5^6\}$ is a reduced set of residues modulo 14.

(b) Show that $\text{hcf}(a, n) = \text{hcf}(n - a, n)$. Use this to prove that the sum of the integers in the reduced set of least positive residues of n is $\dfrac{n\phi(n)}{2}$.

(c) Determine the final two digits of 27^{82} and of 7^{38}.

(d) Show that if $\text{hcf}(a, n) = 1$, then the linear congruence $ax \equiv b \pmod{n}$ has solution $x \equiv ba^{\phi(n)-1} \pmod{n}$. Hence solve the following linear congruences.

(i) $5x \equiv 11 \pmod{12}$

(ii) $3x \equiv 7 \pmod{16}$

(e) If m and n are coprime integers, prove that

$$m^{\phi(n)} + n^{\phi(m)} \equiv 1 \pmod{mn}.$$

(f) Prove that for any given natural number n, there is some positive multiple of n each of whose digits is either 0 or 1.

4 Properties of Euler's ϕ-function

This section explores some more properties of the ϕ-function, with the aim of finding a general formula for $\phi(n)$.

4.1 The multiplicativity of Euler's ϕ-function

We have evaluated $\phi(n)$ for several small values of n and we have discovered the one general result that $\phi(n) = n - 1$ if, and only if, n is prime. Any hope of finding a general formula for $\phi(n)$ in terms of the prime decomposition of n, such as we found for $\sigma(n)$ and $\tau(n)$, would be heightened if it happens that ϕ is a multiplicative function. Fortunately it is.

Theorem 4.1 *Multiplicativity of ϕ*

The function ϕ is multiplicative.

The underlying idea in the proof that we are about to give, showing that $\phi(mn) = \phi(m)\phi(n)$ for coprime integers m and n, is very simple. From a reduced set of residues modulo m and a reduced set of residues modulo n we will construct a set of $\phi(m)\phi(n)$ numbers that will be a reduced set of residues modulo mn. To illustrate this, consider the case $m = 6$ and $n = 7$. Reduced sets of residues for these moduli are $A = \{1, 5\}$ for modulus 6 and $B = \{1, 2, 3, 4, 5, 6\}$ for modulus 7. Now consider the set of $\phi(6)\phi(7) = 12$ integers

$$\{na + mb : a \in A, b \in B\} = \{7a + 6b : a \in A, b \in B\}.$$

This set is

$$\{13, 19, 25, 31, 37, 43, 41, 47, 53, 59, 65, 71\},$$

The first six listed values correspond to $a = 1$ and the second six to $a = 5$.

which, on reducing each number modulo 6×7 and putting the terms into increasing order, is

$$\{1, 5, 11, 13, 17, 19, 23, 25, 29, 31, 37, 41\}.$$

Each of these integers is coprime to 42, and indeed this is a reduced set of residues for modulus 42, showing that $\phi(6 \times 7) = \phi(6)\phi(7)$.

The proof that follows reproduces this construction for general coprime moduli m and n.

Proof of Theorem 4.1 Let m and n be coprime integers; our task is to show $\phi(mn) = \phi(m)\phi(n)$.

Let $\{r_1, r_2, r_3, \ldots, r_{\phi(m)}\}$ and $\{s_1, s_2, s_3, \ldots, s_{\phi(n)}\}$ be reduced sets of residues for the respective moduli m and n. We show that the set of $\phi(m)\phi(n)$ integers

$$\{nr_i + ms_j : 1 \leq i \leq \phi(m), 1 \leq j \leq \phi(n)\}$$

is a reduced set of residues for modulus mn, thereby showing that $\phi(m)\phi(n) = \phi(mn)$. There are three properties to be established:

(a) Each integer in this set is coprime to mn.

(b) No two integers in this set are congruent modulo mn.

(c) Each integer coprime to mn is congruent modulo mn to an integer in this set.

We prove each of these in turn.

(a) Suppose that p is a prime factor of $\text{hcf}(nr_i + ms_j, mn)$. From the fact that p divides mn with $\text{hcf}(m, n) = 1$, we conclude that either p divides m or p divides n but not both. Let us suppose that p divides m. Then, as p divides $nr_i + ms_j$, we must have p divides nr_i and therefore p divides r_i. But then p divides $\text{hcf}(m, r_i) = 1$, which is a contradiction. The alternative supposition that p divides n leads to the similar contradiction that p divides $\text{hcf}(n, s_j)$. No such prime p can exist and so each of the integers $nr_i + ms_j$ is coprime to mn.

Remember that r_i belongs to a reduced set of residues modulo m.

(b) Suppose that $nr + ms \equiv nr' + ms' \pmod{mn}$, where r, r', s and s' are chosen from the reduced sets of residues. Then

$$n(r - r') + m(s - s') = k(mn), \text{ for some integer } k.$$

As m divides two of the terms in this equation it must divide the third, that is, m divides $n(r - r')$. Now $\text{hcf}(m, n) = 1$, and so Euclid's Lemma tells us that m divides $r - r'$, or what amounts to the same thing, $r \equiv r' \pmod{m}$. But r and r' are members of a reduced set of residues modulo m, and so $r = r'$. The same argument shows that $s = s'$ and we conclude that distinct elements of the given set cannot be congruent modulo mn.

(c) Let k be any integer coprime to mn. Since $\text{hcf}(m, n) = 1$, we can write $k = nr + ms$, for some integers r and s. Now suppose there is a prime p that divides both m and r; such a prime would be a common factor of k and mn, contradicting $\text{hcf}(k, mn) = 1$. Hence r is coprime to m and as such is congruent modulo m to some r_i, for $1 \leq i \leq \phi(m)$. And, by exactly the same argument with the roles of m and n interchanged, s is congruent modulo n to one of the s_j in the reduced set. Writing $r = r_i + am$ and $s = s_j + bn$, we have

$$k = nr + ms = nr_i + ms_j + mn(a + b) \equiv nr_i + ms_j \pmod{mn}$$

as required.

See Proposition 4.7 of Book A, Chapter 1.

As (a), (b) and (c) are all established, the proof is complete. ■

Exercise 4.1

Use the multiplicative property of ϕ to evaluate $\phi(24)$, $\phi(70)$ and $\phi(420)$.

4.2 The formula for $\phi(n)$

To obtain a formula that gives the value of $\phi(n)$ in terms of the prime decomposition of n, all we need now is to be able to evaluate ϕ at a general prime power and appeal to the multiplicative property.

Proposition 4.2 *Formula for $\phi(n)$*

If $n > 1$ has prime decomposition $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ then

$$\begin{aligned}\phi(n) &= p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_r}\right) \\ &= n \prod_{\substack{p \mid n \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right).\end{aligned}$$

The formula does not apply for $n = 1$, where $\phi(1) = 1$.

There is one term in the product Π for each prime that divides n, that is, the product is taken over $p_1, p_2, \ldots, p_r$.

Proof To determine $\phi\left(p^k\right)$ we count the number of integers lying between 1 and p^k inclusive that are coprime to p^k. As the only factors of p^k are powers of p, $\text{hcf}\left(m, p^k\right) = 1$ if, and only if, p does not divide m. There are p^{k-1} multiples of p lying between 1 and p^k, namely

$$p, 2p, 3p, \ldots, p^{k-1}p,$$

and the remaining $p^k - p^{k-1}$ integers are coprime to p. Hence,

$$\phi\left(p^k\right) = p^k - p^{k-1} = p^k\left(1 - \frac{1}{p}\right).$$

The multiplicative property of ϕ now gives

$$\begin{aligned}\phi(n) &= \phi\left(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\right) \\ &= \phi\left(p_1^{k_1}\right)\phi\left(p_2^{k_2}\right) \cdots \phi\left(p_r^{k_r}\right) \\ &= p_1^{k_1}\left(1 - \frac{1}{p_1}\right) p_2^{k_2}\left(1 - \frac{1}{p_2}\right) \cdots p_r^{k_r}\left(1 - \frac{1}{p_r}\right) \\ &= p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_r}\right) \\ &= n \prod_{\substack{p \mid n \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right). \quad \blacksquare\end{aligned}$$

With the aid of this formula, computing values of $\phi(n)$ presents little problem. For example,

$$\begin{aligned}\phi(7920) &= \phi\left(2^4\right)\phi\left(3^2\right)\phi(5)\phi(11)\\ &= 2^4\times 3^2\times 5\times 11\times\left(1-\tfrac{1}{2}\right)\left(1-\tfrac{1}{3}\right)\left(1-\tfrac{1}{5}\right)\left(1-\tfrac{1}{11}\right)\\ &= 2^4\times 3^2\times 5\times 11\times\tfrac{1}{2}\times\tfrac{2}{3}\times\tfrac{4}{5}\times\tfrac{10}{11}\\ &= 2^7\times 3\times 5\\ &= 1920.\end{aligned}$$

Exercise 4.2

Determine $\phi(336)$ and $\phi(8p^3)$, where p is any odd prime.

The two examples that follow make use of the multiplicative nature of ϕ and employ the formula for ϕ.

Worked Exercise 4.3

(a) If p is a prime and $n\geq 1$, prove that

$$\phi(pn)=\begin{cases}(p-1)\phi(n), & \text{if } p \text{ does not divide } n,\\ p\phi(n), & \text{if } p \text{ divides } n.\end{cases}$$

Deduce that if d divides n then $\phi(d)$ divides $\phi(n)$.

(b) Find all integers n for which $\phi(n)=10$.

Solution

(a) The case where p does not divide n is easily settled. Here $\text{hcf}(p,n)=1$ and so

$$\phi(pn)=\phi(p)\phi(n)=(p-1)\phi(n).$$

When p does divide n we cannot use the multiplicative property immediately. Let $n=p^km$, where $\text{hcf}(m,p)=1$ and $k\geq 1$. Then

$$\phi(pn)=\phi\left(p^{k+1}m\right)=\phi\left(p^{k+1}\right)\phi(m)=p^{k+1}\left(1-\tfrac{1}{p}\right)\phi(m),$$

whilst

$$p\phi(n)=p\,\phi\left(p^km\right)=p\,\phi\left(p^k\right)\phi(m)=p\times p^k\left(1-\tfrac{1}{p}\right)\phi(m),$$

and the two are seen to be equal.

This result shows that $\phi(n)$ divides $\phi(pn)$ for *any* prime p, regardless of whether or not p is a factor of n. Now if d is a factor of n we may write $n=p_1p_2\cdots p_rd$, where the p_i are (not necessarily distinct) primes that may or may not divide d.

Repeated application of the result gives

$$\phi(d)|\phi(p_1d),\ \phi(p_1d)|\phi(p_1p_2d),\ldots,$$
$$\phi(p_1p_2\cdots p_{r-1}d)|\phi(p_1p_2\cdots p_{r-1}p_rd),$$

and as the last of these terms is $\phi(n)$ we see that $\phi(d)$ divides $\phi(n)$.

(b) We first observe that the formula for $\phi(n)$ can be expressed in a slightly different way that removes the reciprocals of the primes.

Alternative formula for $\phi(n)$

If $n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$, then

$$\phi(n) = p_1^{k_1-1}p_2^{k_2-1}\cdots p_r^{k_r-1}(p_1-1)(p_2-1)\cdots(p_r-1).$$

In this formulation, some (possibly all) of the exponents $k_i - 1$ may be 0.

The main observation to make from this formula is that if p divides n then $p-1$ divides $\phi(n)$. For the case $\phi(n) = 10$, the primes p for which $p-1$ divides 10 are 2, 3 and 11 and so n must be of the form $n = 2^a3^b11^c$ for some non-negative integers a, b and c.

Suppose first that $c = 0$. Then $n = 2^a3^b$ and $\phi(n) = \phi(2^a)\phi(3^b)$. Now $\phi(2^a) = 2^{a-1}$ (for $a \geq 1$) and $\phi(3^b) = 3^{b-1}(3-1)$ (for $b \geq 1$), and as neither of these is divisible by 5 there is no solution. So for $\phi(n) = 10$ we must have $c \geq 1$. In this case we can write $n = 11^cm$, where $\text{hcf}(11, m) = 1$. Then

$$\phi(n) = \phi(11^c)\phi(m) = 11^{c-1} \times 10 \times \phi(m) = 10.$$

This only happens when $c = 1$ and $\phi(m) = 1$. Now $\phi(m) = 1$ for $m = 1$ and $m = 2$ alone (see Exercise 4.3(a) below), and so we have just the two integers $n = 11$ and $n = 22$ that satisfy $\phi(n) = 10$.

Exercise 4.3

(a) Prove that $\phi(n)$ is even for every integer $n > 2$.

(b) Use the formulas obtained for $\sigma(n)$, $\tau(n)$ and $\phi(n)$ to show that if n has prime decomposition $n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$ then:

(i) $\sigma(n)\phi(n) = n^2 \displaystyle\prod_{1\leq i\leq r}\left(1 - \frac{1}{p_i^{k_i+1}}\right)$

(ii) $\tau(n)\phi(n) \geq n$.

Hint: note that $1 - \frac{1}{p} \geq \frac{1}{2}$ for any prime p.

In Proposition 1.5 we discovered that whenever f is a multiplicative function then so too is

$$F(n) = \sum_{d|n} f(d).$$

Which function F will result when we take $f = \phi$? To obtain some feeling for F, let us work out a sample of its values.

$$F(8) = \phi(1) + \phi(2) + \phi(4) + \phi(8) = 1 + 1 + 2 + 4 = 8$$
$$F(9) = \phi(1) + \phi(3) + \phi(9) = 1 + 2 + 6 = 9$$
$$F(10) = \phi(1) + \phi(2) + \phi(5) + \phi(10) = 1 + 1 + 4 + 4 = 10$$

There are no prizes for spotting the pattern!

Proposition 4.4

$$\sum_{d|n} \phi(d) = n$$

If we try applying any of the various formulas concerning the ϕ-function we will struggle to make headway towards proving this result. But there is an elegant, snappy proof, first thought of by Gauss (1777–1855). It hinges on the idea of partitioning the integers from 1 to n inclusive into sets in such a way that two integers belong to the same set if, and only if, they have the same highest common factor with n. For example, for $n = 12$ we have six classes

$$S_1 = \{1, 5, 7, 11\}, \quad S_2 = \{2, 10\}, \quad S_3 = \{3, 9\},$$
$$S_4 = \{4, 8\}, \quad S_6 = \{6\}, \quad S_{12} = \{12\}.$$

The subscript in the label for each class is the highest common factor associated with that class; for example, the two members of S_4 are the positive integers not exceeding 12 whose highest common factor with 12 is equal to 4. The trick now is to form a new set $T_{12/d}$ from S_d by dividing each member of the set by the appropriate highest common factor. This gives

$$T_{12} = \{1, 5, 7, 11\}, \quad T_6 = \{1, 5\}, \quad T_4 = \{1, 3\},$$
$$T_3 = \{1, 2\}, \quad T_2 = \{1\}, \quad T_1 = \{1\}.$$

For example, S_4 becomes $T_{12/4}$ on dividing each element by 4. Notice that each T_i consists of a reduced set of (least positive) residues for modulus i and so the number of elements in T_i is $\phi(i)$. Adding up over the six classes, the sum of the six values of $\phi(i)$, one for each factor i of 12, is equal to the total number of integers in the classes, namely 12.

The proof that follows sets this idea out for a general natural number n.

Proof of Proposition 4.4 Define

$$S_d = \{m : 1 \leq m \leq n \text{ and } \mathrm{hcf}(m, n) = d\}.$$

Now $\mathrm{hcf}(m, n) = d$ if, and only if, $\frac{m}{d}$ and $\frac{n}{d}$ are integers with $\mathrm{hcf}\left(\frac{m}{d}, \frac{n}{d}\right) = 1$. Therefore the number of integers in S_d is equal to the number of positive integers not exceeding $\frac{n}{d}$ that are coprime to $\frac{n}{d}$, that is, $\phi\left(\frac{n}{d}\right)$. As each of the integers $1, 2, \ldots, n$ lies in exactly one class S_d, and there is one class corresponding to each factor d of n,

$$n = \sum_{d|n} \phi\left(\frac{n}{d}\right) = \sum_{d|n} \phi(d). \quad \blacksquare$$

See Exercise 1.2(b) in this chapter.

Exercise 4.4

(a) Find all solutions of $\phi(n) = 18$.

(b) Show that there is no natural number n for which $\phi(n) = 14$. Confirm that $d = 7$ is the smallest natural number for which $\phi(n) = 2d$ has no solutions.

(c) (i) Show that $\phi(n) = \phi(n + 2)$ is satisfied by $n = 2(2p - 1)$ whenever p and $2p - 1$ are odd primes.

(ii) Show that $\phi(n) + 2 = \phi(n + 2)$ is satisfied by $n = 4p$ whenever p and $2p + 1$ are odd primes.

(d) Show that $\phi\left(n^2\right) = n\phi(n)$, for all integers $n \geq 1$. Hence show that if p is a prime and $k \geq 2$, then

$$\phi\Big(\phi\Big(p^k\Big)\Big) = p^{k-2}\phi((p-1)^2).$$

5 Primitive roots

This section extends the idea of primitive roots from Book A, Chapter 4 to include composite moduli.

5.1 Primitive roots generalised

In Exercise 3.4 we saw that the integer 7 had the maximum possible order, $\phi(22) = 10$, modulo 22. When dealing with prime moduli we called an integer with this maximal property a *primitive root*. That definition can be generalised to a composite modulus.

In Subsection 2.2 of Book A, Chapter 4, we investigated primes p that have 10 as a primitive root.

Remember that the condition $\text{hcf}(a, n) = 1$ has to hold for the order of a to be defined.

Definition 5.1 *Primitive root*

If a has order $\phi(n)$ modulo n then a is said to be a **primitive root of** n.

Suppose that a is a primitive root of n and consider the set of integers $\{a, a^2, \ldots, a^{\phi(n)}\}$. Each of the integers in this set is coprime to n because $\text{hcf}(a, n) = 1$. Also, no two members of this set are congruent modulo n because if $a^r \equiv a^s \pmod{n}$, where $1 \leq s \leq r \leq \phi(n)$, then $a^{r-s} \equiv 1 \pmod{n}$ and Proposition 3.5 informs us that the order of a modulo n, namely $\phi(n)$, divides $r - s$; so $r = s$. As the set contains $\phi(n)$ incongruent integers, this set is a reduced set of residues modulo n. In other words, taking the first $\phi(n)$ powers of any primitive root 'generates' a reduced set of residues.

Proposition 5.2

If a is a primitive root of n then

$$\{a, a^2, a^3, \ldots, a^{\phi(n)}\}$$

is a reduced set of residues modulo n.

For example, in Exercise 3.3 we discovered that 3 has order $\phi(14) = 6$ modulo 14, and so is a primitive root of 14. The first six powers of 3 give the members of a reduced set of residues of 14 as follows:

$$3^1 \equiv 3 \pmod{14}, \quad 3^4 \equiv 11 \pmod{14},$$
$$3^2 \equiv 9 \pmod{14}, \quad 3^5 \equiv 5 \pmod{14},$$
$$3^3 \equiv 13 \pmod{14}, \quad 3^6 \equiv 1 \pmod{14}.$$

In Exercise 3.3 we also discovered the order of each of these six numbers modulo 14, as summarised in Table 5.1.

Table 5.1

k	1	2	3	4	5	6
3^k	3	9	13	11	5	1
Order of 3^k	6	3	2	3	6	1

As you would expect from Proposition 3.5, the order of 3^k always divides $\phi(14) = 6$, but notice that this order is 6 when k is coprime to 6, it is 3 when $k = 2$ or 4, it is 2 when $k = 3$, and it is 1 when $k = 6$. This suggests the following result.

Theorem 5.3

If a has order c modulo n then, for any $k \geq 1$, a^k has order

$$\frac{c}{\text{hcf}(c,k)} \quad \text{modulo } n.$$

For example, knowing that 3 has order 6 modulo 14, this theorem tells us that

$$3^2 \text{ has order } \frac{6}{\text{hcf}(6,2)} = 3, \quad 3^4 \text{ has order } \frac{6}{\text{hcf}(6,4)} = 3,$$
$$3^3 \text{ has order } \frac{6}{\text{hcf}(6,3)} = 2, \quad 3^5 \text{ has order } \frac{6}{\text{hcf}(6,5)} = 6.$$

All of these agree with the final row of Table 5.1.

Proof Suppose that a has order c modulo n, and consider a^k. Writing $d = \text{hcf}(c,k)$, let $c = dc'$ and $k = dk'$, where $\text{hcf}(c',k') = 1$. Our task is to show that the order of a^k is c'. Let the order of a^k be r. Then

$$(a^k)^{c'} = (a^{dk'})^{c/d} = a^{k'c} = (a^c)^{k'} \equiv (1)^{k'} \equiv 1 \pmod{n},$$

and so, by Proposition 3.5, c' is a multiple of r.

On the other hand

$$a^{kr} = (a^k)^r \equiv 1 \pmod{n},$$

and so kr is a multiple of c. That is (substituting values for k and c), $dk'r$ is a multiple of dc', which shows that c' divides $k'r$. But then Euclid's Lemma, which we can apply since $\text{hcf}(c',k') = 1$, tells us that c' divides r.

Putting the two parts together, as c' is both a multiple and a factor of r, we have that $c' = r$ as required. ■

As an immediate consequence of this result we can say how many primitive roots n will have and how to find them, provided it has at least one primitive root.

Corollary 5.4

If a is a primitive root of n then a^k is also a primitive root of n if, and only if, k is coprime to $\phi(n)$.

Furthermore, if n has a primitive root, then it has exactly $\phi(\phi(n))$ primitive roots.

Of course, we count only incongruent primitive roots.

Proof Suppose that a is a primitive root of n. Then $\{a, a^2, \ldots, a^{\phi(n)}\}$ is a reduced set of residues for n and so all primitive roots are contained in this set. Now the order of a^k is $\dfrac{\phi(n)}{\text{hcf}(\phi(n), k)}$ and so a^k will be a primitive root precisely when $\text{hcf}(\phi(n), k) = 1$, that is, when k is coprime to $\phi(n)$.

Remember that the primitive roots of n are the integers with order $\phi(n)$.

So the primitive roots are the integers a^k, where k belongs to the set $\{1, 2, \ldots, \phi(n)\}$ and is coprime to $\phi(n)$. By definition of ϕ, there are $\phi(\phi(n))$ such exponents k, and hence $\phi(\phi(n))$ primitive roots of n. ■

Let us put the theorem and its corollary to work.

Worked Exercise 5.5

Given that 2 is a primitive root of 13, find all the primitive roots of 13.

Solution

By Theorem 5.3, 2^k has order 12 when $\text{hcf}(k, 12) = 1$.

As 13 allegedly has one primitive root, there will be in all $\phi(\phi(13)) = \phi(12) = 4$ primitive roots of 13. The $\phi(12)$ natural numbers not exceeding 12 and coprime to it are 1, 5, 7 and 11. According to the corollary, these are the exponents of 2 that give the primitive roots:

$$2^1 = 2;\ 2^5 = 32 \equiv 6;\ 2^7 = 2^5 \times 4 \equiv 11;$$
$$2^{11} = 2^5 \times 2^5 \times 2 \equiv 7 \,(\text{mod } 13),$$

and so the four primitive roots of 13 are 2, 6, 7 and 11.

Table 5.2 shows the primitive roots for all moduli up to 10.

Table 5.2 Primitive roots for small integers

n	$\phi(n)$	$\phi(\phi(n))$	Primitive roots
2	1	1	1
3	2	1	2
4	2	1	3
5	4	2	2, 3
6	2	1	5
7	6	2	3, 5
8	4	2	None
9	6	2	2, 5
10	4	2	3, 7

Note that 8 does not have any primitive roots, as $\phi(8) = 4$ and

$$1^2 \equiv 3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \,(\text{mod } 8)$$

shows that all the integers coprime to 8 have order 1 or 2.

Exercise 5.1

Find all primitive roots, if any, of

(a) 11, and

(b) 15.

5.2 Moduli admitting primitive roots

Having seen that some moduli have primitive roots and some do not, curiosity leads us to ask the question of just which ones do. To round off this chapter we will give the answer, and show where it comes from, but we will omit some of the detail. The key result is the following.

Theorem 5.6

Any integer that can be expressed as the product of two factors that are coprime and each exceeding 2 does not have a primitive root.

Proof Let the integer in question be mn, where $m > 2$, $n > 2$ and $\text{hcf}(m, n) = 1$. Choosing any integer a that is coprime to mn, our goal is to show that $a^k \equiv 1 \pmod{mn}$ for some natural number k less than $\phi(mn)$. In fact $k = \frac{1}{2}\phi(mn)$ will suffice, as we will show.

We first observe that, since $\phi(m)$ and $\phi(n)$ are both even (see Exercise 4.3(a)), it follows that $\frac{1}{2}\phi(m)$ and $\frac{1}{2}\phi(n)$ are integers.

As $\text{hcf}(a, mn) = 1$ we have that $\text{hcf}(a, m) = 1$, from which Euler's Theorem gives $a^{\phi(m)} \equiv 1 \pmod{m}$. Therefore, with $k = \frac{1}{2}\phi(mn)$,

$$a^k = a^{\phi(m)\phi(n)/2} = (a^{\phi(m)})^{\phi(n)/2} \equiv (1)^{\phi(n)/2} \equiv 1 \pmod{m}.$$

The same argument with the roles of m and n interchanged shows $a^k \equiv 1 \pmod{n}$, and putting the two parts together gives $a^k \equiv 1 \pmod{mn}$, as required. ■

Theorem 5.6 may not seem to have made inroads into our immediate problem until you stop and think about which integers cannot be expressed as the product of two coprime factors each exceeding 2. Any number that is divisible by at least two odd primes can be expressed in this way. What does that leave?

Corollary 5.7

If an integer n has a primitive root then it must have one of the following forms.

(a) $n = p^k$, p prime, $k \geq 1$

(b) $n = 2p^k$, p an odd prime, $k \geq 1$

Classes (a) and (b) in the corollary give all the remaining candidates for moduli n that admit a primitive root. But we observe that not all the members of class (a) have the desired property; $8 = 2^3$ has been seen to have no primitive root. In fact, all the powers of 2 from 2^3 onwards do not have primitive roots.

Proposition 5.8

For any $k \geq 3$, 2^k does not have a primitive root.

Proof We know that $\phi\left(2^k\right) = 2^{k-1}$, and the 2^{k-1} least positive residues that are coprime to 2^k are the odd integers up to $2^k - 1$. So our task is to show that for any odd integer a there exists a natural number $r < 2^{k-1}$ such that $a^r \equiv 1 \pmod{2^k}$. It turns out that $r = 2^{k-2}$ has precisely this property for every choice of a. To justify this, let $P(k)$ be the proposition

$a^{2^{k-2}}$ is $a^{(2^{k-2})}$, not $\left(a^2\right)^{k-2}$.

$$a^{2^{k-2}} \equiv 1 \pmod{2^k}.$$

We use mathematical induction to prove that $P(k)$ is true for all $k \geq 3$. The case $P(3)$ is a familiar one: $a^2 \equiv 1 \pmod 8$ for all odd integers a, as seen in the solution to Worked Exercise 3.9. That gives us the basis for induction, and so we turn to the induction step. Assume that $P(h)$ is true for some $h \geq 3$. That is, writing the congruence in a more suitable way,

$$a^{2^{h-2}} = 1 + m2^h, \text{ for some integer } m.$$

Squaring both sides,

$$\begin{aligned}\left(a^{2^{h-2}}\right)^2 &= \left(1 + m2^h\right)^2 \\ &= 1 + 2\left(m2^h\right) + \left(m2^h\right)^2 \\ &= 1 + 2^{h+1}\left(m + m^2 2^{h-1}\right) \\ &\equiv 1 \pmod{2^{h+1}}.\end{aligned}$$

$\left(a^{2^{h-2}}\right)^2 = a^{2^{h-1}}$

Hence $a^{2^{h-1}} \equiv 1 \pmod{2^{h+1}}$, confirming the truth of $P(h+1)$ and completing the induction proof. ■

It turns out that the remaining integers listed in Corollary 5.7 do all have primitive roots and so we have the following result.

Theorem 5.9 *Integers with primitive roots*

The integer n has a primitive root if, and only if, $n = 2$, 4, p^k, $2p^k$, for p an odd prime and $k \geq 1$.

This result, whose proof we delay until Book E, Chapter 18, turns out to be of great importance in algebra, in particular in the study of *fields*.

The first proof of this theorem was given by Gauss in 1801. A key stage of the proof is to show that every odd prime has a primitive root. This part was managed in 1773 by Euler, who proved that every prime p has $\phi(p-1)$ primitive roots.

Exercise 5.2

(a) (i) Which integers have order 6
(1) modulo 17, (2) modulo 18?

(ii) Given that 2 is a primitive root of 19, use Theorem 5.3 to determine which integers have order 6 modulo 19.

(b) Let p be an odd prime and suppose that the integer a has order 3 modulo p. Show that

$$1 + a + a^2 \equiv 0 \,(\text{mod } p).$$

Hence prove that $1 + a$ has order 6 modulo p.

(c) Use Theorem 5.3 to show that if r is a primitive root of an odd prime p and k is even, then r^k is not a primitive root of p. Hence show that if r and s are primitive roots of some odd prime p then rs is not a primitive root of p.

(d) Suppose that the natural number n has a primitive root. Prove that the product of the numbers that form a reduced set of residues of n is congruent modulo n to -1. Is the result necessarily true if n does not have a primitive root?

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should:

(a) Understand the definitions of the τ and σ functions, and be able to use these, together with the factor formulas and the Σ, Π notation, to evaluate expressions involving them.

(b) Understand the two equivalent definitions of multiplicative functions, and be able to apply the given methods to prove when a function is multiplicative.

(c) Understand the term 'perfect number', and the classification of even perfect numbers.

(d) Recall the definition of Mersenne numbers, and understand the stated tests for divisibility of Mersenne numbers.

(e) Understand the statement of Euler's form for odd perfect numbers.

(f) Understand and be able to use the definition of Euler's ϕ-function.

(g) Know the meaning of and be able to find the order of an integer modulo n.

(h) Understand the statement and proof of Euler's Theorem, and its applications to calculating large powers in modular arithmetic and to the factorisation of repunits.

(i) Understand the proof that Euler's ϕ-function is multiplicative, and applications of this result.

(j) Understand the definition of 'primitive root', and the key steps in classifying when an integer has a primitive root.

Solutions to exercises

Solution to Exercise 1.1

n	1	2	3	4	5	6	7	8	9	10
$\tau(n)$	1	2	2	3	2	4	2	4	3	4
$\sigma(n)$	1	3	4	7	6	12	8	15	13	18

Solution to Exercise 1.2

(a) The factors of 8 are 1, 2, 4 and 8 itself. Therefore

$$\begin{aligned}\sum_{d|8}\sigma(d) &= \sigma(1)+\sigma(2)+\sigma(4)+\sigma(8)\\ &= 1+3+7+15 = 26,\end{aligned}$$

and

$$\begin{aligned}\prod_{d|8}\tau(d) &= \tau(1)\times\tau(2)\times\tau(4)\times\tau(8)\\ &= 1\times 2\times 3\times 4 = 24.\end{aligned}$$

(b) (i) If d is a factor of n then $d\times\frac{n}{d}=n$ and so $\frac{n}{d}$ is also a factor of n. Therefore if $d_1, d_2, \ldots, d_r$ are the distinct factors of n then $\frac{n}{d_1}$, $\frac{n}{d_2}, \ldots, \frac{n}{d_r}$ is the same collection of factors, but in a different order. It follows that

$$\begin{aligned}\sum_{d|n} f\left(\frac{n}{d}\right) &= f\left(\frac{n}{d_1}\right)+f\left(\frac{n}{d_2}\right)+\cdots+f\left(\frac{n}{d_r}\right)\\ &= f(d_1)+f(d_2)+\cdots+f(d_r)\\ &= \sum_{d|n} f(d).\end{aligned}$$

(ii) Applying this formula in the case where f is the reciprocal function, we have

$$\sum_{d|n}\frac{1}{d} = \sum_{d|n}\frac{1}{\left(\frac{n}{d}\right)} = \frac{1}{n}\left(\sum_{d|n} d\right) = \frac{\sigma(n)}{n}.$$

Solution to Exercise 1.3

(a) $360 = 2^3\times 3^2\times 5$ and so:

$$\begin{aligned}\tau(360) &= \tau\left(2^3\right)\times\tau\left(3^2\right)\times\tau(5) = 4\times 3\times 2 = 24,\\ \sigma(360) &= \sigma\left(2^3\right)\times\sigma\left(3^2\right)\times\sigma(5)\\ &= (1+2+4+8)(1+3+9)(1+5) = 15\times 13\times 6 = 1170.\end{aligned}$$

(b)
$$\begin{aligned}\tau(2p^3) &= \tau(2)\tau(p^3) = 2\times 4 = 8,\\ \sigma(3p) &= \sigma(3)\sigma(p) = 4(p+1),\\ \sigma(6p^2) &= \sigma(2)\sigma(3)\sigma(p^2) = 3\times 4\times(p^2+p+1) = 12(p^2+p+1).\end{aligned}$$

Solution to Exercise 1.4

From the multiplicative property $f(mn) = f(m)f(n)$, the case $n = 1$ gives $f(m) = f(m)f(1)$. For this to hold we either have $f(1) = 1$ or $f(m) = 0$ for all natural numbers m. In the latter case f is the zero function, which is trivially multiplicative. With the exception of the zero function, all multiplicative functions f have $f(1) = 1$.

Solution to Exercise 1.5

(a) Let m and n be any two integers. If either of these is even, then

$$f(mn) = 0 = f(m)f(n).$$

The condition that $\text{hcf}(m, n) = 1$ is not needed.

On the other hand, if m and n are both odd, then

$$f(mn) = mn = f(m)f(n).$$

Since these cover all cases, it follows that f is a multiplicative function.

$$F(n) = \sum_{d|n} f(d) = \sum_{\substack{d|n \\ d \text{ even}}} f(d) + \sum_{\substack{d|n \\ d \text{ odd}}} f(d) = 0 + \sum_{\substack{d|n \\ d \text{ odd}}} d$$

$$= \text{the sum of the } \textit{odd} \text{ factors of } n.$$

(b) To apply Proposition 1.5 we require a multiplicative function f such that

$$\sigma_r(n) = \sum_{d|n} d^r = \sum_{d|n} f(d).$$

It follows that f must be the function defined by $f(n) = n^r$, and for this function,

$$f(mn) = (mn)^r = m^r n^r = f(m)f(n).$$

So f, and with it σ_r, are multiplicative functions.

Solution to Exercise 1.6

(a) If n is square-free then each exponent in the prime decomposition of n is less than 2; that is, $n = p_1 p_2 \cdots p_r$ is a product of distinct primes and

$$\tau(n) = 2 \times 2 \times 2 \times \cdots \times 2 = 2^r,$$

where r is the number of primes involved.

The converse is not true. For example, $\tau\left(2^3\right) = 4$, and 2^3 is certainly not square-free.

(b) Knowing that $\sigma(n) > n$, one approach is to determine the value of $\sigma(n)$ for each n up to and including 23; all the integers n for which $\sigma(n) = 24$ will turn up in this list. But to be a little more sophisticated, let us make use of the formula

$$\sigma(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}) = (1 + p_1 + p_1^2 + \cdots + p_1^{k_1}) \cdots (1 + p_r + p_r^2 + \cdots + p_r^{k_r}).$$

If n has only one prime factor then

$$\sigma\left(p_1^{k_1}\right) = \left(1 + p_1 + p_1^2 + \cdots + p_1^{k_1}\right) = 24.$$

This is not solvable for $p_1 = 2$ or 3 and when $p_1 \geq 5$ we have $1 + p_1 + p_1^2 > 30$. That leaves $1 + p_1 = 24$, which has solution $p_1 = 23$, and $n = 23$.

If n has more than one prime factor we can appeal to its multiplicative nature, and from $n = rs$ with $r \geq 2$, $s \geq 2$ and $\text{hcf}(r, s) = 1$, we have $\sigma(n) = \sigma(r)\sigma(s) = 24$. Now the least value that $\sigma(r)$ can take is 3 and the only ways that 24 can occur as the product of factors not less than 3 is as 3×8 or 4×6.

The values of $\sigma(n)$ that do not exceed 10 are listed in the solution to Exercise 1.1.

$\sigma(r) = 3$ and $\sigma(s) = 8$ imply $r = 2$ and $s = 7$, giving $n = 14$.

$\sigma(r) = 4$ and $\sigma(s) = 6$ imply $r = 3$ and $s = 5$, giving $n = 15$.

Hence $\sigma(n) = 24$ for $n = 14$, 15 and 23.

(c) From the formula $\tau(n) = (k_1 + 1)(k_2 + 1) \cdots (k_r + 1)$, where the k_i are the exponents in the prime decomposition of n, we see that $\tau(n)$ is odd if, and only if, each exponent is even. Furthermore, each exponent is even if, and only if, n is a square.

Now $\sigma(n)$ is a product of terms $\left(1 + p + p^2 + \cdots + p^k\right)$, one for each prime factor of n. For $\sigma(n)$ to be odd, each of these terms must be odd. When $p = 2$ the term is odd for all values of k. When p is odd $1 + p + p^2 + \cdots + p^k$ is odd only when k is even; so the exponent of each odd prime must be even. We conclude that $\sigma(n)$ is odd when n is any power of 2 (possibly 2^0) times the square of an odd number, that is, $\sigma(n)$ is odd when n is either a square or twice a square.

(d) If d is a factor of n then it has a complementary factor $d' = \frac{n}{d}$ such that $dd' = n$. Now $d \neq d'$ except for the case when n is a square and $d = \sqrt{n}$. So, when n is not a square the $\tau(n)$ factors partition into $\tau(n)/2$ pairs, each pair having product n, and the product of all the factors is $n^{\tau(n)/2}$. On the other hand, when n is a square the $\tau(n)$ factors partition into $(\tau(n) - 1)/2$ pairs, each pair having product n, the remaining factor being $\sqrt{n}$. This time the product of all the factors is equal to $n^{(\tau(n)-1)/2} \times n^{1/2}$, which again is $n^{\tau(n)/2}$.

(e) (i) If n is the prime p, then $n + \sigma(n) = p + (1 + p) = 2p + 1$. For this to be divisible by 3 we require $2p + 1 \equiv 0 \,(\text{mod } 3)$. This has solution $p \equiv 1 \,(\text{mod } 3)$. Advancing to modulus 6,

$$p \equiv 1 \,(\text{mod } 3) \text{ implies } p \equiv 1, 4 \,(\text{mod } 6),$$

but numbers congruent to 4 modulo 6 are divisible by 2 and so cannot be prime. We conclude that, in this case, $p \equiv 1 \,(\text{mod } 6)$.

(ii) If $n = p^2$, for some prime p, then

$$n + \sigma(n) = p^2 + (p^2 + p + 1) = 2p^2 + p + 1.$$

Now when $p = 3$ this quadratic has value 22, which is not divisible by 3. So $p \neq 3$, in which case $p \equiv 1$ or $2 \,(\text{mod } 3)$, and squaring gives $p^2 \equiv 1 \,(\text{mod } 3)$ in each case. It follows that

$$2p^2 + p + 1 \equiv p \not\equiv 0 \,(\text{mod } 3).$$

Hence $n \neq p^2$ if $n + \sigma(n)$ is divisible by 3.

(iii) If $n = pq$ then

$$n + \sigma(n) = pq + (pq + p + q + 1) = 2pq + p + q + 1.$$

Working modulo 3, p can take any of the values 0, 1 or 2 but q can take only values 1 or 2 since $q > 3$ is prime.

The following table gives the possible values modulo 3 of $2pq + p + q + 1$.

	$p \equiv 0$	$p \equiv 1$	$p \equiv 2$
$q \equiv 1$	2	2	2
$q \equiv 2$	0	2	1

The only 0 in the table, corresponding to divisibility by 3, occurs when $p \equiv 0 \,(\mathrm{mod}\ 3)$ and $q \equiv 2 \,(\mathrm{mod}\ 3)$, that is, $p = 3$ and $q \equiv 2 \,(\mathrm{mod}\ 3)$. Now just as in part (a),

$$q \equiv 2 \,(\mathrm{mod}\ 3) \text{ implies } q \equiv 2 \text{ or } 5 \,(\mathrm{mod}\ 6),$$

and as the former value is impossible for an odd prime, we have $q \equiv 5 \,(\mathrm{mod}\ 6)$.

(f) Let m and n be coprime integers. Then

$$\begin{aligned} F(mn) &= f(mn)g(mn), \text{ by the definition of } F, \\ &= (f(m)f(n))\,(g(m)g(n)), \text{ by the multiplicative property of } f \text{ and } g, \\ &= (f(m)g(m))\,(f(n)g(n)), \text{ by rearranging,} \\ &= F(m)F(n). \end{aligned}$$

This confirms that F is multiplicative.

Solution to Exercise 2.1

Suppose that $2^k - 1 = p$, a prime. Then p is odd and

$$\sigma(p) = p + 1 = 2^k.$$

As $\mathrm{hcf}(2^{k-1}, p) = 1$, the multiplicative property of σ gives

$\sigma(2^r) = 2^{r+1} - 1$

$$\begin{aligned} \sigma(n) &= \sigma(2^{k-1}p) = \sigma(2^{k-1})\sigma(p) = (2^k - 1)2^k \\ &= 2(2^{k-1}(2^k - 1)) = 2n. \end{aligned}$$

Thus n is perfect.

Solution to Exercise 2.2

Since q is prime and $\mathrm{hcf}(2, q) = 1$, FLT gives $2^{q-1} \equiv 1 \,(\mathrm{mod}\ q)$, that is,

$$2^{2p} - 1 \equiv 0 \,(\mathrm{mod}\ q).$$

Factorising the left-hand side of this congruence gives

$$(2^p - 1)(2^p + 1) \equiv 0 \,(\mathrm{mod}\ q).$$

As q is prime, Euclid's Lemma gives

$$q|(2^p - 1) \text{ or } q|(2^p + 1),$$

which is the claimed result as $M_p = 2^p - 1$.

Solution to Exercise 2.3

The square root of M_{17} is approximately $2^{17/2}$, which is close to 362. So if M_{17} is composite, it must have a prime factor of the form $34k+1$ that is less than 362. Of the ten numbers of this form that are less than 362, namely

35, 69, **103**, **137**, 171, 205, **239**, 273, **307**, 341,

just the four emboldened numbers are prime. Either M_{17} is prime or it is divisible by one of these four primes.

$$2^{17/2} = 2^8\sqrt{2} = 256 \times \sqrt{2} \\ \simeq 362.04$$

Solution to Exercise 2.4

(a) Suppose that n is perfect. Then $\sum_{d|n} d = 2n$. Using Exercise 1.2(b),

$$2n = \sum_{d|n} d = \sum_{d|n} \frac{n}{d} = n\left(\sum_{d|n} \frac{1}{d}\right),$$

from which $\sum_{d|n} \frac{1}{d} = 2$ follows.

The argument reverses. If $\sum_{d|n} \frac{1}{d} = 2$ then $\sum_{d|n} \frac{n}{d} = 2n$, from which we conclude that n is perfect.

If $d < n$ is a factor of n then $\sum_{r|d} \frac{1}{r} < \sum_{r|n} \frac{1}{r}$, since each factor of d is also a factor of n but at least one factor of n (namely n itself) is not a factor of d. If n is perfect, the sum on the right of the inequality is equal to 2 and so the sum on the left must be less than 2. Hence d is not perfect.

(b) (i) Working modulo 6, the successive powers of 2 alternate 2, 4, 2, 4, ..., with value 2 when the exponent is odd. As k is odd,

$$n = 2^{k-1}\big(2^k - 1\big) \equiv 4(2-1) \equiv 4 \,(\text{mod } 6).$$

(ii) Working modulo 9, the successive powers of 2 cycle 2, 4, 8, 7, 5, 1, 2, 4, 8, ... As k is odd, 2^{k-1} is congruent modulo 9 to one of 4, 7 or 1 giving, respectively, the following possibilities for n:

$$n = 2^{k-1}\big(2^k - 1\big) \equiv 4(8-1) = 28 \equiv 1 \,(\text{mod } 9),$$
$$n = 2^{k-1}\big(2^k - 1\big) \equiv 7(5-1) = 28 \equiv 1 \,(\text{mod } 9),$$
$$n = 2^{k-1}\big(2^k - 1\big) \equiv 1(2-1) \equiv 1 \,(\text{mod } 9).$$

In all cases n is congruent modulo 9 to 1.

(iii) $n = \frac{1}{2} \times 2^k(2^k - 1) = \frac{1}{2} \times m(m+1)$, where $m = 2^k - 1$. This is the triangular number T_m.

(c) (i) $\sigma(24) = 60 > 48$, and so 24 is abundant.
$\sigma(25) = 31 < 50$, and so 25 is deficient.
$\sigma(26) = 42 < 52$, and so 26 is deficient.
$\sigma(27) = 40 < 54$, and so 27 is deficient.
$\sigma(28) = 56 = 56$, and so 28 is perfect.
$\sigma(29) = 30 < 58$, and so 29 is deficient.
$\sigma(30) = 72 > 60$, and so 30 is abundant.

(ii) $\sigma(p^r) = 1 + p + p^2 + \cdots + p^{r-1} + p^r < 2p^r$ since

$$1 + p + p^2 + \cdots + p^{r-1} = \frac{p^r - 1}{p - 1} < p^r.$$

The required inequality can also be proved using mathematical induction.

So p^r is deficient.

(iii) Let p and q be distinct primes with $p < q$. Then

$$\sigma(pq) = \sigma(p)\sigma(q) = (p + 1)(q + 1) = pq + p + q + 1.$$

This will be less than $2pq$ whenever $p + q + 1 < pq$. Now

$$\begin{aligned} p + q + 1 &\leq 2q \quad \text{(equality holding only when } q = p + 1; \\ &\qquad\qquad \text{that is, when } p = 2, q = 3) \\ &\leq pq \quad \text{(equality holding only when } p = 2). \end{aligned}$$

Hence any number of the form pq is deficient except for the single case $pq = 6$, which is perfect.

(iv) The smallest odd abundant number is $n = 3^3 \times 5 \times 7$, which has $\sigma(n) = 40 \times 6 \times 8 = 1920 > 1890 = 2n$.

(v) $\sigma(mn) = \sigma(m)\sigma(n) > 2mn$, since $\sigma(m) > 2m$ and $\sigma(n) \geq n$.

(vi) Since $\text{hcf}(2^{k-1}, 2^k - 1) = 1$,

$$\sigma\big(2^{k-1}(2^k - 1)\big) = \sigma\big(2^{k-1}\big)\,\sigma\big(2^k - 1\big) = \big(2^k - 1\big)\,\sigma\big(2^k - 1\big).$$

Now, for any $m > 1$, $\sigma(m) \geq m + 1$, with equality holding only when m is prime. Hence, if $2^k - 1$ is not prime, then $\sigma\big(2^k - 1\big) > 2^k$, and

$$\sigma\big(2^{k-1}(2^k - 1)\big) > 2 \times 2^{k-1}\big(2^k - 1\big).$$

This shows that $2^{k-1}\big(2^k - 1\big)$ is abundant.

(d) (i) $\sigma(220) = \sigma\big(2^2 \times 5 \times 11\big) = 7 \times 6 \times 12 = 504 = 220 + 284$,
$\sigma(284) = \sigma\big(2^2 \times 71\big) = 7 \times 72 = 504 = 220 + 284$.

$\sigma\big(2^4 \times 23 \times 47\big) = 31 \times 24 \times 48 = 35\,712 = 17\,296 + 18\,416$,
$\sigma\big(2^4 \times 1151\big) = 31 \times 1152 = 35\,712 = 17\,296 + 18\,416$.

$\sigma\big(2^7 \times 191 \times 383\big) = 255 \times 192 \times 384 = 18\,800\,640$,
$\sigma\big(2^7 \times 73\,727\big) = 255 \times 73\,728 = 18\,800\,640$,
$2^7 \times 191 \times 383 + 2^7 \times 73\,727 = 9\,363\,584 + 9\,437\,056 = 18\,800\,640$.

So in each case the pair is amicable.

(ii) $\sigma(p) = p + 1$, and so p can be amicable only with 1. But $\sigma(1) = 1 \neq p + 1$.

Solution to Exercise 3.1

n	Reduced set of residues	$\phi(n)$
1	{1}	1
2	{1}	1
3	{1,2}	2
4	{1,3}	2
5	{1,2,3,4}	4
6	{1,5}	2
7	{1,2,3,4,5,6}	6
8	{1,3,5,7}	4
9	{1,2,4,5,7,8}	6
10	{1,3,7,9}	4

0 does not appear in a reduced set of residues modulo $n = 1$ because $0 \equiv 1 \pmod 1$.

Solution to Exercise 3.2

Of the integers in the set

$$\{\mathbf{1}, 2, \mathbf{3}, 4, \mathbf{5}, 6, 7, 8, \mathbf{9}, 10, \mathbf{11}, 12, \mathbf{13}\}$$

only the six emboldened integers are coprime to 14, and so $\phi(14) = 6$.

$3^6 = (3^3)^2 = (27)^2 \equiv (-1)^2 = 1 \pmod{14}$
$5^6 = (5^2)^3 = 25^3 \equiv (-3)^3 = -27 \equiv 1 \pmod{14}$

Solution to Exercise 3.3

The least positive residues in a reduced set modulo 14 are 1, 3, 5, 9, 11 and 13.

$1^1 \equiv 1 \pmod{14}$, and so 1 has order 1.
$3^2 \equiv 9$, $3^3 \equiv 13$, $3^4 \equiv 11$, $3^5 \equiv 5$, $3^6 \equiv 1 \pmod{14}$, and so 3 has order 6.
$5^2 \equiv 11$, $5^3 \equiv 13$, $5^4 \equiv 9$, $5^5 \equiv 3$, $5^6 \equiv 1 \pmod{14}$, and so 5 has order 6.
$9^2 \equiv 11$, $9^3 \equiv 1 \pmod{14}$, and so 9 has order 3.
$11^2 \equiv 9$, $11^3 \equiv 1 \pmod{14}$, and so 11 has order 3.
$13^2 \equiv (-1)^2 \equiv 1 \pmod{14}$, and so 13 has order 2.

Solution to Exercise 3.4

(a) The prime decomposition of 21 is 3×7, and so $\phi(21) = 12$, as can be seen by counting how many integers in $\{1, 2, \ldots, 19, 20\}$ are not divisible by either 3 or 7. So the possible orders of 4 modulo 21 are 1, 2, 3, 4, 6 and 12.

$$4^1 \equiv 4; \quad 4^2 \equiv 16; \quad 4^3 = 64 \equiv 1 \pmod{21},$$

and so 4 has order 3.

(b) $\phi(22) = 10$ and so the possible orders of 7 modulo 22 are 1, 2, 5 and 10.

$7^1 \equiv 7; \quad 7^2 = 49 \equiv 5;$
$7^5 = 7^2 \times 7^2 \times 7 \equiv 5 \times 5 \times 7 \equiv 3 \times 7 \equiv 21 \pmod{22}.$

As 7 does not have one of the smaller orders it must have order 10. Of course, $7^{10} \equiv 1 \pmod{22}$ by Euler's Theorem.

Solution to Exercise 3.5

As $100 = 2^2 5^2$, $\phi(100)$ is equal to the number of integers in the set $\{1, 2, 3, \ldots, 99, 100\}$ that are not multiples of either 2 or 5, that is, those numbers that end in 1, 3, 7 or 9. There are 40 such numbers and so $\phi(100) = 40$.

The final two digits of a number are determined by its least positive residue modulo 100.

(a) By Euler's Theorem, $17^{40} \equiv 1 \pmod{100}$. Therefore

$$17^{83} = (17^{40})^2 17^3 \equiv 17^3 = 289 \times 17 \equiv (-11) \times 17 = -187 \equiv 13 \pmod{100},$$

and so the final two digits of 17^{83} are 13.

(b) Since $19^{40} \equiv 1 \pmod{100}$, 19^{39} is a solution of $19x \equiv 1 \pmod{100}$.

We can simplify this congruence by multiplying through by any integer that is coprime to 100:

$$\begin{aligned} 19x \equiv 1 \pmod{100} &\iff 209x \equiv 11 \pmod{100} \\ &\iff 9x \equiv 11 \pmod{100}. \end{aligned}$$

We can then replace the right-hand side with a congruent number, thus taking advantage of the result that a number is divisible by 9 when the sum of its digits is divisible by 9:

$$\begin{aligned} 19x \equiv 1 \pmod{100} &\iff 9x \equiv 711 \pmod{100} \\ &\iff x \equiv 79 \pmod{100}, \end{aligned}$$

and the final two digits of 19^{39} are 79.

Solution to Exercise 3.6

(a) Let $P(m)$ be the statement that R_{mn} is divisible by R_n.

$P(1)$ is certainly true as it asserts that R_n is divisible by R_n, and so we have the basis for induction.

Turning to the induction step, suppose that $P(k)$ is true and consider $R_{(k+1)n}$.

$$R_{(k+1)n} = R_{kn+n} = R_{kn} \times 10^n + R_n.$$

By the induction hypothesis, R_n divides R_{kn} and so divides both terms on the right-hand side of this equation. It follows that R_n divides the left-hand side, $R_{(k+1)n}$, as required.

Hence by induction, R_{mn} is divisible by R_n for all integers $m \geq 1$.

(b) (i) Euler's Theorem gives $10^{p-1} \equiv 1 \pmod{p}$. In other words p divides $10^{p-1} - 1$, which amounts to p divides $9R_{p-1}$. However, $\mathrm{hcf}(9, p) = 1$, and so p divides R_{p-1}.

Remember that $p > 5$. The result is not true for $p = 2$, $p = 3$ or $p = 5$.

(ii) Suppose that n does not divide $p - 1$. Then $p - 1 = nq + r$, where $0 < r < n$.

We are given that p divides R_n (and therefore divides R_{nq} by part (a)). Moreover, we know that p divides $R_{p-1} = R_{nq+r}$ by part (b)(i). If we now look at

$$R_{nq+r} = R_{nq} \times 10^r + R_r,$$

we see that p divides two of the terms in this equation and so must divide the third. That is, p divides R_r, contradicting the fact that R_n is the smallest repunit which is a multiple of p. Thus $p - 1$ is a multiple of n, as required.

(iii) (1) We know that 7 divides R_6, and by part (b)(ii) the only smaller repunits that may be divisible by 7 are $R_2 = 11$ and $R_3 = 111$. As neither of these is divisible by 7, R_6 is the smallest such repunit.

(2) 13 divides R_{12}, and the candidates for smaller multiples of 13 are R_2, R_3, R_4 and R_6. In fact, 13 divides $R_6 = 111\,111$ but not R_2, R_3 or R_4. So the smallest repunit divisible by 13 is R_6.

Solution to Exercise 3.7

The modulus is the product of coprime factors 3, 5, 16 and 17. For the primes 3, 5 and 17, FLT gives, respectively, $n^3 \equiv n \pmod 3$, $n^5 \equiv n \pmod 5$ and $n^{17} \equiv n \pmod{17}$, for all integers n. Therefore

$$n^{33} = (n^3)^{11} \equiv n^{11} = (n^3)^3 n^2 \equiv n^3 n^2 \equiv n^3 \equiv n \pmod 3,$$
$$n^{33} = (n^5)^6 n^3 \equiv n^6 n^3 = n^5 n^4 \equiv nn^4 = n^5 \equiv n \pmod 5,$$
$$n^{33} = n^{17} n^{16} \equiv nn^{16} = n^{17} \equiv n \pmod{17}.$$

Note that FLT must be used in this form because the result is needed for all odd integers; so cases with n congruent to 0 must not be overlooked. Handling the 15 via Euler's Theorem, with $n^{\phi(15)} \equiv 1 \pmod{15}$, will not cater for all cases.

For modulus 16 we must use Euler's Theorem. As every odd integer is coprime to 16 while every even integer is not, it follows that $\phi(16) = 8$. Therefore, for all odd integers n, $n^8 \equiv 1 \pmod{16}$ and so

$$n^{33} = n(n^8)^4 \equiv n(1)^4 \equiv n \pmod{16}.$$

All four congruences hold for any odd integer n and so, for such n, $n^{33} \equiv n \pmod{15 \times 16 \times 17}$ follows.

Solution to Exercise 3.8

(a) (i) Those integers between 1 and 20 inclusive that are coprime to 20 form the reduced set of residues $\{1, 3, 7, 9, 11, 13, 17, 19\}$.

(ii) As $\phi(14) = 6$ (see Exercise 3.2) and $\text{hcf}(5, 14) = 1$, the set $\{5, 5^2, 5^3, 5^4, 5^5, 5^6\}$ will be a reduced set of residues provided they are distinct modulo 14. This is the case since

$$5^2 \equiv 11,\ 5^3 \equiv 5 \times 11 \equiv 13,\ 5^4 \equiv 5 \times 13 \equiv 9,$$
$$5^5 \equiv 5 \times 9 \equiv 3,\ 5^6 \equiv 5 \times 3 \equiv 1.$$

(b) In establishing the Euclidean Algorithm in Subsection 5.1 of Book A, Chapter 1 we had the result

$$\text{hcf}(a, b) = \text{hcf}(a - qb, b), \quad \text{for any integer } q.$$

In Book A, Chapter 1 we used r in place of $a - qb$.

Applying this here:

$$\begin{aligned}\mathrm{hcf}(n-a,n) &= \mathrm{hcf}(n-a-n,n)\\ &= \mathrm{hcf}(-a,n) = \mathrm{hcf}(a,n).\end{aligned}$$

The result is readily verified for the case $n = 2$ and so we will assume that $n \geq 3$. Consider the set of integers from 1 to $n-1$ inclusive that are coprime to n. By the above, these break into pairs, a and $n-a$, which have sum n.

If $a = n - a$ then $2a = n$. So a is a factor of n and $\mathrm{hcf}(a,n) = a = \frac{n}{2} > 1$, and so a is not coprime to n, which is a contradiction.

Hence the $\phi(n)$ integers break into $\phi(n)/2$ pairs of distinct integers with sum n, and so the total sum is $n\phi(n)/2$.

(c) The final two digits of an integer are given by its least positive residue modulo 100. Now $\phi(100) = 40$ (see Exercise 3.5) and so $27^{40} \equiv 1 \pmod{100}$. Therefore

$$27^{82} = (27^{40})^2 27^2 \equiv 27^2 = 729 \equiv 29 \pmod{100},$$

and so the final two digits are 29.

Since $7^{40} \equiv 1 \pmod{100}$, 7^{38} is a solution of $7^2 x \equiv 1 \pmod{100}$. However,

$$\begin{aligned}7^2x \equiv 1 \pmod{100} &\iff 49x \equiv 1 \pmod{100}\\ &\iff 49x \equiv 301 \pmod{100}\\ &\iff 7x \equiv 43 \pmod{100}\\ &\iff 7x \equiv 343 \pmod{100}\\ &\iff x \equiv 49 \pmod{100},\end{aligned}$$

and so the required two digits are 49.

(d) If $x \equiv ba^{\phi(n)-1} \pmod{n}$ then $ax \equiv ba^{\phi(n)} \pmod{n}$, and since Euler's Theorem gives $a^{\phi(n)} \equiv 1 \pmod{n}$, we have $ax \equiv b \pmod{n}$. As $\mathrm{hcf}(a,n) = 1$ this is the unique solution modulo n, by Theorem 3.4 of Book A, Chapter 3.

Applying this result to the given congruences and noting that $\phi(12) = 4$ and $\phi(16) = 8$:

(i) $5x \equiv 11 \pmod{12}$ has solution
$x \equiv 11 \times 5^3 = 55 \times 25 \equiv 55 \equiv 7 \pmod{12}$; and

(ii) $3x \equiv 7 \pmod{16}$ has solution
$x \equiv 7 \times 3^7 = 7 \times 81 \times 27 \equiv 7 \times 1 \times 11 \equiv 13 \pmod{16}$.

(e) If m and n are coprime then, by Euler's Theorem, $m^{\phi(n)} \equiv 1 \pmod{n}$ and $n^{\phi(m)} \equiv 1 \pmod{m}$. Therefore

$$\begin{aligned}&m^{\phi(n)} + n^{\phi(m)} \equiv 0 + 1 = 1 \pmod{m}, \text{ and}\\ &m^{\phi(n)} + n^{\phi(m)} \equiv 1 + 0 = 1 \pmod{n}.\end{aligned}$$

As $\text{hcf}(m,n)=1$, it follows from Corollary 1.7 of Book A, Chapter 3 that $m^{\phi(n)}+n^{\phi(m)}\equiv 1 \pmod{mn}$.

(f) Suppose that $n=2^r5^sm$, where either of r and s may be 0 and $\text{hcf}(m,10)=1$. We have seen that there is some multiple of m, say km, that is a repunit; that is, all the digits of km are 1s. Hence, if we can find a multiple of $kn=2^r5^skm$ that equalises the exponents of 2 and 5, so that we are multiplying km by some power of 10, then the resulting number will be the repunit km with a string of 0s appended.

Let $t=\max\{r,s\}$. Then $(2^{t-r}5^{t-s}k)\,n=10^tkm$ is a suitable multiple.

Solution to Exercise 4.1

$\phi(24)=\phi(3\times 8)=\phi(3)\phi(8)=2\times 4=8$

$\phi(70)=\phi(2\times 5\times 7)=\phi(2)\phi(5)\phi(7)=1\times 4\times 6=24$

$\phi(420)=\phi(3\times 4\times 5\times 7)=\phi(3)\phi(4)\phi(5)\phi(7)=2\times 2\times 4\times 6=96$

Solution to Exercise 4.2

$$\begin{aligned}\phi(336)&=\phi(2^4)\phi(3)\phi(7)\\&=2^4\times 3\times 7\times \tfrac{1}{2}\times\tfrac{2}{3}\times\tfrac{6}{7}=96\end{aligned}$$

$$\phi(8p^3)=\phi(8)\phi(p^3)=4p^3\left(1-\tfrac{1}{p}\right)=4p^2(p-1)$$

Solution to Exercise 4.3

(a) If n is divisible by an odd prime p then $\phi(n)$ is divisible by $p-1$ and consequently is even. If n is not divisible by an odd prime then $n=2^k$ and $\phi(n)=2^{k-1}$, which is even for $k>1$. The only numbers not covered are $n=1$ and $n=2$; in both these cases $\phi(n)=1$.

(b) (i) From the formulas for σ and ϕ:

$$\begin{aligned}\sigma(n)\phi(n)&=\frac{p_1^{k_1+1}-1}{p_1-1}\cdots\frac{p_r^{k_r+1}-1}{p_r-1}\times p_1^{k_1-1}\cdots p_r^{k_r-1}(p_1-1)\cdots(p_r-1)\\&=(p_1^{k_1+1}-1)\cdots(p_r^{k_r+1}-1)p_1^{k_1-1}\cdots p_r^{k_r-1}\\&=p_1^{2k_1}\cdots p_r^{2k_r}\left(1-\frac{1}{p_1^{k_1+1}}\right)\cdots\left(1-\frac{1}{p_r^{k_r+1}}\right)\\&=n^2\prod_{1\le i\le r}\left(1-\frac{1}{p_i^{k_i+1}}\right).\end{aligned}$$

(ii) Since $\left(1-\frac{1}{p}\right)\ge\frac{1}{2}$ for each prime p, we note that

$$\phi(n)=n\prod_{1\le i\le r}\left(1-\frac{1}{p_i}\right)\ge\frac{n}{2^r}.$$

Now $\tau(n)=(k_1+1)(k_2+1)\cdots(k_r+1)\ge 2^r$, since each $k_i\ge 1$.

From the two inequalities, $\phi(n)\tau(n)\ge n$ follows.

Solution to Exercise 4.4

(a) If a prime p is a factor of n then $p-1$ is necessarily a factor of $\phi(n)$. The primes p for which $p-1$ is a factor of 18 are 2, 3, 7 and 19.

Suppose that 19 divides n. Clearly 19^2 cannot divide n because $\phi\big(19^2\big)$ is well in excess of 18. So $n = 19m$, where $\text{hcf}(19, m) = 1$ and

$$\phi(n) = \phi(19)\phi(m) = 18\phi(m) = 18.$$

Now $\phi(m) = 1$ only for $m = 1$ and 2; so we obtain two solutions, $n = 19$ and $n = 38$.

Suppose next that the largest prime factor of n is 7. Again, 7^2 cannot be a factor of n because $\phi\big(7^2\big) = 42$. So $n = 7m$, where $\text{hcf}(7, m) = 1$ and

$$\phi(n) = \phi(7)\phi(m) = 6\phi(m) = 18.$$

This is impossible because we have seen that $\phi(m)$ is even for all $m > 2$ and $\phi(m) = 1$ for $m = 1$ or 2.

The remaining possibilities are $n = 2^r$, $n = 3^s$ and $n = 2^r 3^s$, for positive integers r and s.

For $n = 2^r$,

$$\phi(2^r) = 2^{r-1} \neq 18.$$

For $n = 3^s$,

$$\phi(3^s) = 3^s \times \tfrac{2}{3} = 2 \times 3^{s-1} = 18,$$

giving the solution $n = 3^3 = 27$.

For $n = 2^r 3^s$,

$$\phi(2^r 3^s) = 2^r 3^s \times \tfrac{1}{2} \times \tfrac{2}{3} = 2^r 3^{s-1} = 18,$$

giving the solution $n = 2^1 3^3 = 54$.

So there are four solutions, namely $n = 19$, 27, 38 and 54.

(b) By the same argument as employed in the solution to part (a) above, if $\phi(n) = 14$ then the only possible prime factors of n are 2 and 3. It follows that n must have one of the prime decompositions $n = 2^r$, $n = 3^s$ or $n = 2^r 3^s$. The corresponding values of the ϕ-function are 2^{r-1}, $2 \times 3^{s-1}$ and $2^r 3^{s-1}$. As none of these has a factor 7, the value 14 cannot arise.

All smaller even numbers arise as values of ϕ:

$$\phi(3) = 2,\ \phi(5) = 4,\ \phi(7) = 6,$$
$$\phi(16) = 8,\ \phi(11) = 10,\ \phi(13) = 12.$$

(c) (i) If p and $2p-1$ are odd primes then, for $n = 2(2p-1)$,

$$\phi(n) = \phi(2)\phi(2p-1) = 1 \times (2p-2),$$
$$\phi(n+2) = \phi(4p) = \phi(4)\phi(p) = 2(p-1),$$

and the two are seen to be equal.

(ii) If $n = 4p$, where p and $2p + 1$ are odd primes, then

$$\phi(n) + 2 = \phi(4p) + 2 = 2(p-1) + 2 = 2p,$$
$$\phi(n+2) = \phi(2(2p+1)) = \phi(2)\phi(2p+1) = 2p,$$

and again the two are equal.

(d) The key here is the observation that the set of primes that divide n^2 is exactly the same as the set that divide n. Hence

$$\phi(n^2) = n^2 \prod_{p|n} \left(1 - \frac{1}{p}\right) \text{ and } \phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

from which $\phi(n^2) = n\phi(n)$ follows immediately.

$$\begin{aligned}\phi(\phi(p^k)) &= \phi(p^{k-1}(p-1)) = \phi(p^{k-1})\phi(p-1), \quad \text{as } \operatorname{hcf}(p^{k-1}, p-1) = 1,\\ &= p^{k-2}(p-1)\phi(p-1) = p^{k-2}\phi((p-1)^2),\end{aligned}$$

the final step following from $\phi(n^2) = n\phi(n)$ with $n = p - 1$.

Solution to Exercise 5.1

(a) As $\phi(\phi(11)) = \phi(10) = 4$, there are either none or four primitive roots of 11. In fact 2 is seen to be a primitive root, since its powers modulo 11 begin

$$2,\ 2^2 = 4,\ 2^3 = 8,\ 2^4 \equiv 5 \text{ and } 2^5 \equiv 10 \pmod{11}.$$

Therefore the order of 2 exceeds 5, and as it divides $\phi(11) = 10$, it must actually be 10.

The four primitive roots of 11 are therefore 2^1, 2^3, 2^7 and 2^9, where the exponents $\{1, 3, 7, 9\}$ form a reduced set of residues for $\phi(11) = 10$. Simplifying these values modulo 11 we get

$$2,\ 8,\ 2^7 = 2^4 \times 2^3 \equiv 5 \times 8 \equiv 7 \text{ and } 2^9 = 2^5 \times 2^4 \equiv 10 \times 5 \equiv 6.$$

(b) Here $\phi(\phi(15)) = \phi(8) = 4$ and so there are either no primitive roots of 15 or four of them. The reduced set of least positive residues modulo 15 is $\{1, 2, 4, 7, 8, 11, 13, 14\}$ and the primitive roots – that is, integers of order 8 – are to be found in this set.

Now $2^4 \equiv 1 \pmod{15}$ and so 2 is not a primitive root as it has order 4. Similarly, $4^2 \equiv 1 \pmod{15}$, $14^2 \equiv (-1)^2 \equiv 1 \pmod{15}$ and $11^2 \equiv (-4)^2 \equiv 1 \pmod{15}$ show that none of 4, 11 or 14 is a primitive root. As 1 always has order 1 we are left with just three candidates for primitive roots, namely 7, 8 and 13. As there has to be either none or four in all, we conclude that 15 has no primitive roots.

Solution to Exercise 5.2

(a) (i) (1) Since $\phi(17) = 16$ and 6 does not divide 16, there are no integers of order 6 modulo 17.

(2) $\phi(18) = 6$ and so there may be integers of order 6 modulo 18. The candidates are those numbers in a reduced set of residues of 18, $\{1, 5, 7, 11, 13, 17\}$. Checking successive powers of these numbers modulo 18 reveals that they have respective orders 1, 6, 3, 6, 3 and 2, the elements having order 6 being 5 and 11.

(ii) $\phi(19) = 18$ and as 6 divides 18 there may be elements of order 6 modulo 19. Given that 2 is a primitive root of 19, Theorem 5.3 tells us that 2^k has order $\dfrac{18}{\text{hcf}(k, 18)}$. This will be equal to 6 when $\text{hcf}(18, k) = 3$, that is, for $k = 3$ and 15. The numbers that have order 6 modulo 19 are therefore $2^3 = 8$ and $2^{15} \equiv 12$.

(b) Since $a^3 \equiv 1 \pmod{p}$ we have

$$a^3 - 1 = (a-1)\big(1 + a + a^2\big) \equiv 0 \pmod{p}.$$

Now a is not congruent modulo p to 1 for, if so, a would have order 1. Hence, appealing to Euclid's Lemma, $1 + a + a^2 \equiv 0 \pmod{p}$.

From this congruence we observe that $1 + a \equiv -a^2 \pmod{p}$ and so the successive powers of $1 + a$ modulo p are

$$-a^2, a^4, -a^6, a^8, -a^{10}, a^{12}, \ldots.$$

Utilising the fact that $a^3 \equiv 1 \pmod{p}$, these simplify to

$$-a^2, a, -1, a^2, -a, 1, \ldots,$$

showing that $1 + a$ has order 6.

(c) If r is a primitive root of the odd prime p then r has order $p - 1$ modulo p and, from Theorem 5.3, the primitive roots of p are the powers r^k for those natural numbers k with $\text{hcf}(k, p-1) = 1$. As $p - 1$ is even we see immediately that k must be odd.

Let s be another primitive root of p. Then $s = r^k$, for some odd k. But now $rs = r^{k+1}$ cannot be a primitive root since $k + 1$ is even.

(d) If r is a primitive root of n then its powers generate the reduced set of residues $\left\{r, r^2, \ldots, r^{\phi(n)}\right\}$. Note in particular that $r^{\phi(n)} \equiv 1 \pmod{n}$, and so from the reduced set of residues we can form pairs $\left\{r^k, r^{\phi(n)-k}\right\}$ that have the property $r^k r^{\phi(n)-k} \equiv 1 \pmod{n}$. This pairing accounts for all the integers in the reduced set except two, namely the element $r^{\phi(n)/2}$, which would be paired with itself, and the final element $r^{\phi(n)}$. As $r^{\phi(n)} \equiv 1 \pmod{n}$, it follows that the product of all the elements in the reduced set of residues is congruent to $r^{\phi(n)/2} \pmod{n}$. It remains to observe that $r^{\phi(n)/2} \equiv -1 \pmod{n}$; this holds because $\text{hcf}(n, -1) = 1$ and so -1 must feature in the reduced set of residues, and $(-1)x \equiv 1 \pmod{n}$ has unique solution $x \equiv -1 \pmod{n}$, confirming that -1 is the second element paired with itself.

This argument assumes that $\phi(n)$ is even, which is the case for $n > 2$. The result is trivially true for $n = 2$.

The result is not necessarily true if n does not have a primitive root. For example, consider $n = 8$. The product of the four integers in the reduced set of least positive residues is $1 \times 3 \times 5 \times 7 \equiv 1 \pmod{8}$.

Chapter 10

Quadratic reciprocity

Introduction

In Book A, Chapter 4 we studied polynomial congruences. We found that, provided we take congruence to a prime modulus, a polynomial congruence of degree k has at most k incongruent solutions. In this respect polynomial congruences are similar to polynomial equations. In this chapter we are going to concentrate on polynomial congruences of degree 2, called **quadratic congruences**. It turns out that the key problem is to solve the congruence

In most cases the modulus will be an odd prime.

$$x^2 \equiv a \pmod{p}.$$

In other words, we will try to find which integers are squares modulo p. In Book A and in the last chapter we solved such problems for small primes p. For instance, by working out 0^2, 1^2, 2^2, 3^2 and 4^2 modulo 5 we see that the only squares modulo 5 are 0, 1 and 4. It follows that the quadratic congruences

$$x^2 \equiv 2 \pmod{5} \text{ and } x^2 \equiv 3 \pmod{5}$$

are each unsolvable.

As the task of solving quadratic equations presents little challenge, the thought of devoting a whole chapter to quadratic congruences might, understandably, seem unexciting. Be patient! Investigation of when a quadratic congruence has solutions (without actually solving it) leads us through three classic results of number theory: Euler's Criterion, Gauss's Lemma and the Law of Quadratic Reciprocity. The last is a most fascinating result, described by Gauss as the 'gem of higher arithmetic'. The consequences of these results are far-reaching in other areas of number theory. In this chapter we will use them to establish several more special cases of Dirichlet's Theorem on primes in arithmetic progressions, which we first investigated in Book A, Chapter 2.

Chapter guide

This chapter has four sections, of which the first two form the basis for the results in the last two. The final section, on the Law of Quadratic Reciprocity, is somewhat longer than the others, and builds on the results of Section 3 as well as the first two sections.

1 Euler's Criterion

This chapter begins by stating the problem of solving quadratic congruences, and establishing the first of the three results – Euler's Criterion.

1.1 Quadratic residues

Consider the quadratic congruence

$$ax^2 + bx + c \equiv 0 \pmod{p},$$

$a \not\equiv 0 \pmod{p}$
$\iff \text{hcf}(a, p) = 1$
$\iff p$ does not divide a.

where p is an odd prime and $a \not\equiv 0 \pmod{p}$. In an attempt to solve this congruence, we can write the left-hand side in an alternative form by 'completing the square' to get all involvement of x into a single squared term. To achieve that, we first multiply through by $4a$, which will not affect the solutions since $\text{hcf}(4a, p) = 1$:

$$4a^2x^2 + 4abx + 4ac \equiv 0 \pmod{p}.$$

This is essentially the process used to obtain the 'quadratic formula'

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

That is,

$$(2ax + b)^2 + 4ac - b^2 \equiv 0 \pmod{p},$$

and adding $b^2 - 4ac$ to both sides,

$$(2ax + b)^2 \equiv b^2 - 4ac \pmod{p}.$$

If we now let $d = b^2 - 4ac$ and put $y = 2ax + b$, the quadratic congruence takes the simplified form

$$y^2 \equiv d \pmod{p}.$$

For each solution y_0 of this congruence we can recover a solution of the original congruence by solving $2ax + b \equiv y_0 \pmod{p}$, which is known to be uniquely solvable since $\text{hcf}(2a, p) = 1$.

To illustrate this, let us look at a specific worked exercise.

Worked Exercise 1.1

Solve $2x^2 + 2x + 1 \equiv 0 \pmod{29}$.

Solution

Begin by multiplying through by $4 \times 2 = 8$,

$$16x^2 + 16x + 8 \equiv 0 \pmod{29},$$

and then complete the square,

$$(4x + 2)^2 + 4 \equiv 0 \pmod{29}.$$

Now add -4 to both sides:

$$(4x+2)^2 \equiv -4 \pmod{29}.$$

Putting $y = 4x+2$, and replacing -4 by the congruent integer 25, gives

$$y^2 \equiv 25 \pmod{29},$$

which has the two solutions

$$y \equiv 5 \pmod{29} \quad \text{and} \quad y \equiv -5 \pmod{29}.$$

Lagrange's Theorem (Book A, Chapter 4, Theorem 4.3) guarantees that there are no more than these two.

The solutions in x are therefore obtained by solving two linear congruences:

$$\begin{aligned} 4x+2 \equiv 5 \pmod{29} &\iff 4x \equiv 3 \equiv 32 \pmod{29} \\ &\iff x \equiv 8 \pmod{29}, \\ 4x+2 \equiv -5 \pmod{29} &\iff 4x \equiv -7 \equiv -36 \pmod{29} \\ &\iff x \equiv -9 \equiv 20 \pmod{29}. \end{aligned}$$

So the two solutions are $x \equiv 8, 20 \pmod{29}$.

The solutions obtained can, of course, be checked by substituting 8 and 20 back into the original congruence.

Exercise 1.1

Solve the following quadratic congruences using the process illustrated in Worked Exercise 1.1.

(a) $5x^2 + 7x + 1 \equiv 0 \pmod{13}$

(b) $x^2 + x + 1 \equiv 0 \pmod{5}$

(c) $x^2 + 3x + 4 \equiv 0 \pmod{7}$

The examples in Exercise 1.1 show that a quadratic congruence to prime modulus may have 0, 1 or 2 solutions. They also pinpoint what is the key issue in solving $ax^2 + bx + c \equiv 0 \pmod{p}$. By multiplying by $4a$ and completing the square we get the following result.

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. The congruence $ax^2 + bx + c \equiv 0 \pmod{p}$ has solution(s) if, and only if, $b^2 - 4ac$ is congruent modulo p to a square.

The number $b^2 - 4ac$ is called the **discriminant** of the quadratic $ax^2 + bx + c$.

What we need is a way of determining which integers are the squares modulo p. First we introduce some new terminology.

Definition 1.2 *Quadratic residues*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. If the congruence $x^2 \equiv a \pmod{p}$ has a solution then a is said to be a **quadratic residue** of p. Otherwise a is a **quadratic non-residue** of p. We refer to whether a is a quadratic residue or quadratic non-residue of p as being the **quadratic character** of a modulo p.

Notice that if $a \equiv b \pmod{p}$ then the congruence $x^2 \equiv a \pmod{p}$ has a solution if, and only if, the congruence $x^2 \equiv b \pmod{p}$ has a solution. Thus congruent integers are of the same quadratic character, that is, both are quadratic residues of p or both are quadratic non-residues of p. For this reason it will be enough to consider the quadratic character of the non-zero least positive residues modulo p. So when, as in the coming worked exercise, we ask for the quadratic residues or non-residues of p, the expected answer will be a collection of integers in the range from 1 to $p-1$.

Worked Exercise 1.3

Find the quadratic residues and the quadratic non-residues of 11.

Solution

To list the quadratic residues of 11, it suffices to determine the squares $1^2, 2^2, \ldots, 10^2$ modulo 11.

$$1^2 \equiv 1 \pmod{11}$$
$$2^2 \equiv 4 \pmod{11}$$
$$3^2 \equiv 9 \pmod{11}$$
$$4^2 \equiv 5 \pmod{11}$$
$$5^2 \equiv 3 \pmod{11}$$
$$6^2 \equiv (-5)^2 \equiv 3 \pmod{11}$$
$$7^2 \equiv (-4)^2 \equiv 5 \pmod{11}$$
$$8^2 \equiv (-3)^2 \equiv 9 \pmod{11}$$
$$9^2 \equiv (-2)^2 \equiv 4 \pmod{11}$$
$$10^2 \equiv (-1)^2 \equiv 1 \pmod{11}$$

Therefore the quadratic residues of 11 are 1, 3, 4, 5 and 9. The quadratic non-residues of 11 are the remaining non-zero least positive residues, namely 2, 6, 7, 8 and 10.

Exercise 1.2

Determine the quadratic residues and the quadratic non-residues of 13. Hence determine the values of c for which the congruence $x^2 + x - c \equiv 0 \pmod{13}$ has solutions.

You may have noticed that our definition excludes 0 as a quadratic residue despite the fact that 0 is certainly a square. We also removed from consideration the case $p = 2$, where the only non-zero residue, 1, is a square. The major reason is that these decisions enable us to express results in neater forms. For example, the congruence

$$x^2 \equiv a \pmod{p}$$

has no solutions if a is a quadratic non-residue of p, one solution if $a \equiv 0 \pmod{p}$ and two solutions if a is a quadratic residue of p. In addition, with these definitions, as we have seen for the examples $p = 11$ and $p = 13$, there are the same number of quadratic residues as there are quadratic non-residues. We establish this fact now.

If $x \equiv b \pmod{p}$ is a solution then so too is $x \equiv -b \pmod{p}$, and $b \not\equiv -b \pmod{p}$ except when $b \equiv 0 \pmod{p}$.

Proposition 1.4 *The quadratic residues of p*

For any odd prime p there are $\frac{p-1}{2}$ quadratic residues and $\frac{p-1}{2}$ quadratic non-residues. The quadratic residues are congruent modulo p to the integers $1^2, 2^2, 3^2, \ldots, \left(\frac{p-1}{2}\right)^2$.

Proof The quadratic residues of p are the integers that result from the evaluation of the squares $1^2, 2^2, 3^2, \ldots, (p-1)^2$ modulo p. But $r^2 = (-r)^2$ and so these $p - 1$ integers fall into congruent pairs modulo p, namely

$$1^2 \equiv (p-1)^2,\ 2^2 \equiv (p-2)^2,\ \ldots,\ \left(\tfrac{p-1}{2}\right)^2 \equiv \left(\tfrac{p+1}{2}\right)^2.$$

Notice that we require p to be odd here.
As $r^2 \not\equiv 0 \pmod{p}$, for $1 \le r < p$, the integer 0 is not listed amongst these.

Therefore each quadratic residue of p is congruent to one of the $\frac{p-1}{2}$ integers $1^2, 2^2, 3^2, \ldots, \left(\frac{p-1}{2}\right)^2$.

To complete the proof it remains to show that no two of these integers are congruent modulo p. With that objective in mind, suppose that

$$r^2 \equiv s^2 \pmod{p}, \text{ for some } 1 \le s \le r \le \tfrac{p-1}{2}.$$

Our goal is to deduce from this that $r = s$. Now $r^2 \equiv s^2 \pmod{p}$ means that p is a factor of $r^2 - s^2$ or, after factorising, p divides $(r - s)(r + s)$. Euclid's Lemma takes us a step further: either p divides $r - s$ or p divides $r + s$. The latter is impossible since $2 \le r + s \le p - 1$. As to the former, the fact that $r - s$ lies in the range $0 \le r - s < \frac{p-1}{2}$ confirms that p can divide $r - s$ only when $r - s = 0$; that is, when $r = s$.

Thus there are exactly $\frac{p-1}{2}$ quadratic residues and hence $\frac{p-1}{2}$ quadratic non-residues. ■

1.2 Euler's Criterion

We will keep returning to the task of actually solving congruences, but for the time being we concentrate on the key question that confronts us when deciding whether or not a quadratic congruence has solutions:

Given that $a \not\equiv 0 \pmod{p}$, is a a quadratic residue of p?

Proposition 1.4 suggests one line of attack by providing a means of listing all the quadratic residues of p. However, the amount of computation it entails, even for a relatively small prime p, forces us to look around for some alternative method. Our next result, first attributed to Euler, provides such an alternative method. Whilst it can still involve cumbersome computations, it is a very useful theoretical result for which we will find numerous applications.

Theorem 1.5 *Euler's Criterion*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. Then a is a quadratic residue of p if, and only if, $a^{(p-1)/2} \equiv 1 \pmod{p}$ and is a quadratic non-residue of p if, and only if, $a^{(p-1)/2} \equiv -1 \pmod{p}$.

The proof that we are going to give of Euler's Criterion involves careful pairing of the integers forming the set of non-zero least positive residues modulo p. First we illustrate the line of proof for the case of the prime 13.

We take the slightly easier case of a quadratic non-residue first.

We have seen in Exercise 1.2 that 2 is a quadratic non-residue of 13. Consider the congruence $bx \equiv 2 \pmod{13}$, where b is any of the integers 1, 2, ..., 12. Since $\text{hcf}(b, 13) = 1$ this congruence has a unique solution. The solution is certainly not $x \equiv 0 \pmod{13}$ and neither is it $x \equiv b \pmod{13}$, for then $b^2 \equiv 2 \pmod{13}$, which would contradict the fact that 2 is a quadratic non-residue. Hence, for each choice of b, the solution of $bx \equiv 2 \pmod{13}$ is one of the *other* integers in the set $\{1, 2, \ldots, 12\}$. It follows that these 12 integers break into pairs whose product is congruent modulo 13 to 2. These pairs are as follows.

$$1 \times 2 \equiv 2 \pmod{13} \qquad 3 \times 5 \equiv 2 \pmod{13}$$
$$4 \times 7 \equiv 2 \pmod{13} \qquad 6 \times 9 \equiv 2 \pmod{13}$$
$$8 \times 10 \equiv 2 \pmod{13} \qquad 11 \times 12 \equiv 2 \pmod{13}$$

Multiplying these six congruences together, we obtain

$$12! \equiv 2^6 \pmod{13}.$$

But we know from Wilson's Theorem (Book A, Chapter 4, Theorem 3.1) that $12! \equiv -1 \pmod{13}$, and so

$$2^6 \equiv -1 \pmod{13},$$

which is the claim of Euler's Criterion for the case of the quadratic non-residue 2 of 13.

There is nothing special here about the prime 13, nor about the chosen integer 2. The crux of the general proof that follows shortly is to show that, for any odd prime p and any quadratic non-residue a of p, the integers in the set $\{1, 2, \ldots, p-1\}$ break into pairs with product congruent modulo p to a.

We discovered that 3 is a quadratic residue of 13 in Exercise 1.2.

To illustrate Euler's Criterion for the case of a quadratic residue we proceed in a similar vein. Take $a = 3$ and consider solutions of $bx \equiv 3 \pmod{13}$. The difference this time is that the congruence $x^2 \equiv 3 \pmod{13}$ does have solutions, namely $x \equiv 4 \pmod{13}$ and $x \equiv 9 \pmod{13}$. The remaining ten integers in $\{1, 2, \ldots, 12\}$ break into pairs whose product is congruent modulo 13 to 3.

$$1 \times 3 \equiv 3 \pmod{13} \qquad 2 \times 8 \equiv 3 \pmod{13}$$
$$5 \times 11 \equiv 3 \pmod{13} \qquad 6 \times 7 \equiv 3 \pmod{13}$$
$$10 \times 12 \equiv 3 \pmod{13}$$

If we multiply together the left-hand sides of these congruences the outcome will not be 12!, since the integers 4 and 9 are missing. To rectify this, so that we can subsequently appeal to Wilson's Theorem, we include the additional congruence whose left-hand side is 4×9:

$$4 \times 9 \equiv 36 \equiv -3 \pmod{13}.$$

(As you will see shortly, it is no accident that the right-hand side of this congruence has turned out to be the negative of the quadratic residue under consideration.) Multiplying all six congruences together, we reach

$$12! \equiv 3^5 \times (-3) \equiv -3^6 \pmod{13},$$

whereupon invocation of Wilson's Theorem leads to $3^6 \equiv 1 \pmod{13}$. This is what Euler's Criterion claims for the quadratic residue 3 of 13.

Our task now is to convert these illustrations into general arguments.

Proof of Euler's Criterion (Theorem 1.5) Let a be a quadratic non-residue of p and let b be any integer in $\{1, 2, \ldots, p-1\}$. The linear congruence $bx \equiv a \pmod{p}$ has a unique solution b' also in $\{1, 2, \ldots, p-1\}$. Note that $b \neq b'$, for otherwise $b^2 \equiv a \pmod{p}$ and a would be a quadratic residue of p. It follows that the integers in $\{1, 2, \ldots, p-1\}$ fall into $\frac{p-1}{2}$ pairs b, b' with the property that $bb' \equiv a \pmod{p}$. Hence

$$(p-1)! = 1 \times 2 \times 3 \times \cdots \times (p-1) \equiv a \times a \times \cdots \times a \equiv a^{(p-1)/2} \pmod{p}.$$

If we now bring in Wilson's Theorem, $(p-1)! \equiv -1 \pmod{p}$, we reach

$$a^{(p-1)/2} \equiv -1 \pmod{p}, \text{ for any quadratic non-residue } a.$$

Let us now attempt the same argument for the case of a being a quadratic residue of p. This time the congruence $x^2 \equiv a \pmod{p}$ has a solution. In fact it will have exactly two solutions. Lagrange's Theorem guarantees that it has at most two solutions, and if c is one solution in $\{1, 2, \ldots, p-1\}$ then $p - c$ is certainly another since

$$(p-c)^2 \equiv (-c)^2 \equiv c^2 \equiv a \pmod{p},$$

and $c \neq p - c$ as $p = 2c$ is impossible.

If we remove c and $p-c$ from $\{1,2,\ldots,p-1\}$ then, as before, the remaining integers fall into $\frac{p-3}{2}$ pairs b, b' with the property that $bb' \equiv a \pmod{p}$. So this time

$$\begin{aligned}(p-1)! &= 1 \times 2 \times \cdots \times c \times \cdots \times (p-c) \times \cdots \times (p-1)\\ &\equiv a \times a \times \cdots \times a \times c \times (p-c) \pmod{p}\\ &\equiv a^{(p-3)/2}(-c^2) \pmod{p}\\ &\equiv a^{(p-3)/2}(-a) \pmod{p}\\ &\equiv -a^{(p-1)/2} \pmod{p}.\end{aligned}$$

Replacing $(p-1)!$ by -1 and cancelling by -1 leads to

$$a^{(p-1)/2} \equiv 1 \pmod{p}, \text{ for any quadratic residue } a.$$

We have now shown that if a is a quadratic residue, then $a^{(p-1)/2} \equiv 1 \pmod{p}$ and if a is a quadratic non-residue, then $a^{(p-1)/2} \equiv -1 \pmod{p}$.

To complete the 'only if' parts, suppose first that $a^{(p-1)/2} \equiv 1 \pmod{p}$. As $a \not\equiv 0 \pmod{p}$, it is either a quadratic residue or a quadratic non-residue. If we assume that a is a non-residue then $a^{(p-1)/2} \equiv -1 \pmod{p}$ and this would imply that $1 \equiv -1 \pmod{p}$, which is false for an odd prime. So if $a^{(p-1)/2} \equiv 1 \pmod{p}$ then a is a quadratic residue. The non-residue case follows similarly. ■

1.3 Uses of Euler's Criterion

We have already determined the quadratic residues of the prime 13 but there is no harm in tackling the same problem by a totally different method. As you will see, Euler's Criterion is very useful for determining whether or not a given integer is a quadratic residue, but it is less helpful when we wish to determine all the quadratic residues of a given prime.

Worked Exercise 1.6

Use Euler's Criterion to determine the quadratic residues and non-residues of the prime 13.

Solution

To determine whether or not a is a quadratic residue of 13, Euler's Criterion tells us to investigate $a^{(13-1)/2} = a^6$ in the certain knowledge that for any a in the set $\{1, 2, 3, \ldots, 12\}$ either $a^6 \equiv 1 \pmod{13}$, in which case a is a quadratic residue, or $a^6 \equiv -1 \pmod{13}$, in which case a is a quadratic non-residue.

$$1^6 \equiv 1 \pmod{13}$$
$$2^6 = 16 \times 4 \equiv 12 \equiv -1 \pmod{13}$$
$$3^6 = 3^3 \times 3^3 \equiv 1 \times 1 \equiv 1 \pmod{13}$$
$$4^6 = (4^2)^3 \equiv 3^3 \equiv 1 \pmod{13}$$
$$5^6 = (5^2)^3 \equiv (-1)^3 \equiv -1 \pmod{13}$$
$$6^6 = (6^2)^3 \equiv (-3)^3 \equiv -1 \pmod{13}$$

No further computation is needed because

$$(13 - r)^6 \equiv (-r)^6 \equiv r^6 \pmod{13},$$

and so

$$7^6 \equiv 6^6, 8^6 \equiv 5^6, 9^6 \equiv 4^6, 10^6 \equiv 3^6, 11^6 \equiv 2^6 \text{ and } 12^6 \equiv 1^6 \pmod{13}.$$

The list of quadratic residues of 13 is therefore 1, 3, 4, 9, 10 and 12; the non-residues are the other six non-zero least positive residues, namely 2, 5, 6, 7, 8 and 11.

Exercise 1.3

Determine the quadratic character of 5 for each of the primes 11, 17 and 19.

Primitive roots provide an alternative way of determining the quadratic residues of an odd prime p. Suppose that r is a primitive root of p. Then the set of integers $\{r, r^2, \ldots, r^{p-1}\}$ is a reduced set of residues for p. That is, after replacing each by its least positive residue modulo p (and reordering as appropriate), it is the set $\{1, 2, \ldots, p-1\}$. Half of these integers are quadratic residues and the other half are quadratic non-residues, their designations being readily seen: the $\frac{p-1}{2}$ integers $\{r^2, r^4, \ldots, r^{p-1}\}$, being even powers of the primitive root and hence squares, are the quadratic residues.

In Chapter 9 we mentioned (without proof) that every prime does have a primitive root.

Worked Exercise 1.7

Given that 2 is a primitive root of 19, determine the quadratic residues of 19.

Solution

The first $\frac{19-1}{2} = 9$ even powers of 2 give the quadratic residues of 19 whilst the first 9 odd powers of 2 give the quadratic non-residues.

In each case we multiply the previous value by 4 and reduce modulo 19.

$$2^2 = 4$$
$$2^4 = 16$$
$$2^6 = 64 \equiv 7 \pmod{19}$$
$$2^8 \equiv 7 \times 4 = 28 \equiv 9 \pmod{19}$$
$$2^{10} \equiv 9 \times 4 = 36 \equiv 17 \pmod{19}$$
$$2^{12} \equiv 17 \times 4 = 68 \equiv 11 \pmod{19}$$
$$2^{14} \equiv 11 \times 4 = 44 \equiv 6 \pmod{19}$$
$$2^{16} \equiv 6 \times 4 = 24 \equiv 5 \pmod{19}$$
$$2^{18} \equiv 5 \times 4 \equiv 1 \pmod{19}$$

The quadratic residues of 19 are therefore 1, 4, 5, 6, 7, 9, 11, 16 and 17.

From the identity

$$(ab)^{(p-1)/2} = a^{(p-1)/2} b^{(p-1)/2}$$

and knowledge of the quadratic character of a and b, Euler's Criterion tells us immediately whether or not the product ab is a quadratic residue. For example, if a is a quadratic residue of p, so that $a^{(p-1)/2} \equiv 1 \pmod{p}$, and b is a quadratic non-residue of p, so that $b^{(p-1)/2} \equiv -1 \pmod{p}$, then multiplying the two together gives $(ab)^{(p-1)/2} \equiv -1 \pmod{p}$, showing that ab is a quadratic non-residue. Looking similarly at the other three combinations we obtain Table 1.1, which is essentially determined by the products of 1 (for residue) and -1 (for non-residue).

Table 1.1 Table of residue multiplication

	residue	non-residue
residue	residue	non-residue
non-residue	non-residue	residue

You may wish to refer to this table in answering the following exercises.

Exercise 1.4

Let p be an odd prime.

(a) If $b \not\equiv 0 \pmod{p}$ and the congruence $ax^2 \equiv b \pmod{p}$ has a solution, what can you deduce about the quadratic characters of a and b?

(b) If -1 is a quadratic residue of p, what can you infer about p?

(c) If a is a quadratic residue of p, under what circumstances is $p - a$ also a quadratic residue?

Exercise 1.5

(a) Solve the following congruences.

(i) $2x^2 + 4x + 1 \equiv 0 \pmod{7}$

(ii) $3x^2 + x - 3 \equiv 0 \pmod{11}$

(b) For which integers c does the congruence $2x^2 - 5x + c \equiv 0 \pmod{13}$ have solutions? Solve the congruence for the case $c = 1$.

The quadratic residues of 13 are found in Worked Exercise 1.6.

(c) Use Euler's Criterion to determine whether or not 3 is a quadratic residue of the following primes.

(i) 23 (ii) 29

(d) Quadratic residues can be defined for composite $n > 2$ by analogy with the prime case. Each integer a satisfying $\mathrm{hcf}(a, n) = 1$ is either a quadratic residue or non-residue of n depending on whether or not the congruence $x^2 \equiv a \pmod{n}$ has solutions.

Prove that if a is a quadratic residue of n then $a^{\phi(n)/2} \equiv 1 \pmod{n}$.

Remember that $\phi(n)$ is even for $n > 2$ and so $\phi(n)/2$ is an integer.

Show that the converse of this result is false by exhibiting coprime integers a and n for which $a^{\phi(n)/2} \equiv 1 \pmod{n}$ but for which $x^2 \equiv a \pmod{n}$ has no solutions.

2 The Legendre symbol

In order to proceed with our investigation of quadratic congruences, in this section we introduce and develop some notation.

2.1 Properties of the Legendre symbol

In the preceding work we have many times made the statement 'a is a quadratic (non-)residue of p'. There is a convenient shorthand notation for such statements. It was first introduced by Legendre (1752–1833) in the course of an incomplete proof of the Law of Quadratic Reciprocity. Naturally enough the notation is named after him.

You can learn more about Legendre in the *History Reader* for this topic.

In other texts the notations $\left(\frac{a}{p}\right)$ and $(a|p)$ are sometimes used for the Legendre symbol. Remember that (a/p) does not mean, in any sense, 'a divided by p'.

We read (a/p) as 'a slash p'.

Definition 2.1 *Legendre symbol*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. The **Legendre symbol** (a/p) is defined as follows:

$$(a/p) = \begin{cases} 1, & \text{if } a \text{ is a quadratic residue of } p, \\ -1, & \text{if } a \text{ is a quadratic non-residue of } p. \end{cases}$$

As an illustration of the Legendre symbol we can summarise the results of Worked Exercise 1.6, which determined the quadratic residues and quadratic non-residues of 13, by writing

$$(1/13) = (3/13) = (4/13) = (9/13) = (10/13) = (12/13) = 1;$$
$$(2/13) = (5/13) = (6/13) = (7/13) = (8/13) = (11/13) = -1.$$

The following theorem contains no essentially new information; it simply collates facts discovered about quadratic residues in the previous section and re-expresses them in terms of the Legendre symbol.

Theorem 2.2 *Properties of the Legendre symbol*

Let p be an odd prime and let $a \not\equiv 0 \pmod{p}$ and $b \not\equiv 0 \pmod{p}$. Then the following properties hold.

(a) If $a \equiv b \pmod{p}$ then $(a/p) = (b/p)$.

(b) $(a^2/p) = 1$

(c) $(ab/p) = (a/p)(b/p)$

(d) $(a/p) \equiv a^{(p-1)/2} \pmod{p}$

(e) $(-1/p) = \begin{cases} 1, & \text{if } p \equiv 1 \pmod{4}, \\ -1, & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

Note that the exponent, $(p-1)/2$, in property (d) is a fraction, *not* a Legendre symbol.

Property (a) says that congruent integers have the same quadratic character. Property (b) makes the obvious claim that any square is a quadratic residue. Property (c) expresses, in succinct form, the multiplicative property of Table 1.1. Property (d) is, of course, Euler's Criterion; when applied to the particular case $a = -1$, it gives the result of property (e), which we have included because we will refer frequently to this important property.

Exercise 2.1

At each stage of the following evaluation of $(30/11)$, state which of the properties of Theorem 2.2 is being used.

$$\begin{aligned}(30/11) &= (8/11), \quad \text{by ?}\\ &= (2/11)(4/11), \quad \text{by ?}\\ &= (2/11), \quad \text{by ?}\\ &= (-9/11), \quad \text{by ?}\\ &= (-1/11)(9/11), \quad \text{by ?}\\ &= (-1/11), \quad \text{by ?}\\ &= -1, \quad \text{by ?}\end{aligned}$$

2.2 Applications using the Legendre symbol

We are going to encounter many exercises involving quadratic residues. Rather than falling back on first principles, we will solve them through manipulation of the properties in Theorem 2.2. We begin with two worked exercises illustrating how the Legendre symbol can be used.

Worked Exercise 2.3

(a) Does the congruence $x^2 + 6x + 11 \equiv 0 \,(\text{mod } 29)$ have a solution?

(b) Find all the quadratic residues of 37.

Solution

(a) As the discriminant is $6^2 - 4 \times 11 = -8$, we must decide whether or not -8 is a quadratic residue of 29, and so we evaluate the Legendre symbol $(-8/29)$.

$$\begin{aligned}(-8/29) &= (-2/29)(4/29), \quad \text{by property (c)},\\ &= (-2/29), \quad \text{since } (4/29) = 1, \text{ by property (b)},\\ &= (-1/29)(2/29), \quad \text{by property (c)},\\ &= (2/29), \quad \text{since } (-1/29) = 1, \text{ by property (e)},\\ &\equiv 2^{(29-1)/2} = 2^{14}, \quad \text{by property (d)},\\ &= 2^5 \times 2^5 \times 2^4 \equiv 3 \times 3 \times 16\\ &\equiv -1 \,(\text{mod } 29).\end{aligned}$$

As the discriminant is a quadratic non-residue we conclude that the congruence does not have a solution.

The properties referred to are those of Theorem 2.2, and congruence is to modulus 29 throughout.

(b) We know that 37 has 18 quadratic residues and we could go about finding them by determining the least positive residue modulo 37 of each of 1^2, 2^2, 3^2, ..., 18^2, as we did for the prime 11 in Worked Exercise 1.3. Instead we will look for some shortcuts.

For a start, we know from property (b) that 1, 4, 9, 16, 25 and 36 are quadratic residues. As $36 \equiv -1 \pmod{37}$, property (a) gives $(-1/37) = 1$, a fact that we could have established alternatively from property (e). Using this together with first property (c) and then property (a) gives

$$1 = (-1/37)(4/37) = (-4/37) = (33/37),$$

showing that 33 is a quadratic residue. In just the same way, $-9 \equiv 28$, $-16 \equiv 21$ and $-25 \equiv 12$ are quadratic residues of 37.

Having found that 12 and 28 are quadratic residues we can 'remove the square 4' from each, using first property (c) and then property (b), to get

$$1 = (12/37) = (3/37)(4/37) = (3/37)$$

and

$$1 = (28/37) = (7/37)(4/37) = (7/37).$$

So 3 and 7 are quadratic residues and, again using $(-1/37) = 1$, we find that $-3 \equiv 34$ and $-7 \equiv 30$ are also quadratic residues.

We have now found a total of 14 quadratic residues, and so there are 4 still to be found. There are many ways of discovering them. For example, reversing the previous procedure and 'multiplying by the square 4' before using property (a) gives

$$1 = (4/37)(12/37) = (48/37) = (11/37)$$

and

$$1 = (4/37)(16/37) = (64/37) = (27/37),$$

which shows that 11 and 27, and with them their negatives, $-11 \equiv 26$ and $-27 \equiv 10$, are quadratic residues.

The list of quadratic residues of 37 is now complete:

$$1, 3, 4, 7, 9, 10, 11, 12, 16, 21, 25, 26, 27, 28, 30, 33, 34 \text{ and } 36.$$

Exercise 2.2

(a) Using only the properties of Theorem 2.2, evaluate the following Legendre symbols.

(i) $(3/11)$ (ii) $(2/17)$ (iii) $(40/11)$

(b) Show that $(a/31) = 1$ if, and only if, $(2a/31) = 1$. Use this fact to obtain all quadratic residues of 31.

Exercise 2.3

(a) Making use of properties of Theorem 2.2, determine whether or not the following congruences have solutions.

(i) $x^2 + 8x + 4 \equiv 0 \pmod{13}$

(ii) $x^2 + 7x + 9 \equiv 0 \pmod{23}$

(b) Evaluate the following Legendre symbols.

(i) $(18/19)$ (ii) $(45/19)$ (iii) $(2/19)$

(c) Confirm that $(-1/29) = 1$ and hence that ± 1, ± 4, ± 9 and ± 16 are quadratic residues of 29. Determine the remaining quadratic residues of 29.

(d) If p is an odd prime and $\text{hcf}(a, p) = 1$ deduce, using Euler's Criterion and Wilson's Theorem, that

$$(p-1)! \equiv -(a/p)a^{(p-1)/2} \pmod{p}.$$

3 Gauss's Lemma

Our second result on the solution of quadratic congruences is Gauss's Lemma, and this section contains its proof and applications.

3.1 Proof of Gauss's Lemma

One of our objectives in this chapter is to prove, and use, the classical theorem known as the Law of Quadratic Reciprocity. This result will help us overcome the computational difficulties of determining the quadratic character of an integer, even for a large integer and large prime modulus. The fascination that this result has had for mathematicians over the years is exemplified by the fact that over 100 proofs of the Law have been published (seven of them due to Gauss himself – he was not finding further proofs for their own sake, but rather trying to find a method of proof that would generalise from the quadratic case to higher powers). The majority of the published proofs hinge on a subsidiary result, Gauss's Lemma, which is the subject matter of this section. Gauss's Lemma offers yet another method of determining the quadratic character of an integer modulo an odd prime and, like Euler's Criterion, while not proving to be a very practicable test when large integers are involved, it is a useful theoretical result for which further applications arise.

Theorem 3.1 *Gauss's Lemma*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. Let S be the set

$$S = \left\{a, 2a, 3a, \ldots, \tfrac{p-1}{2}a\right\}$$

consisting of the first $\frac{p-1}{2}$ positive multiples of a. If n denotes the number of members of S whose least positive residue modulo p exceeds $\frac{p}{2}$, then $(a/p) = (-1)^n$.

Before setting about the proof, we look at an example that should serve to clarify the statement of Gauss's Lemma and, more importantly, will illustrate the line to be adopted in the general proof.

Let $p = 19$ and $a = 11$. The set S in Gauss's Lemma consists of the first $\frac{19-1}{2} = 9$ positive multiples of 11, namely

$$S = \{11, 22, 33, 44, 55, 66, 77, 88, 99\}.$$

Replacing each member of S by its remainder when it is divided by 19, that is, by the least positive residue to which it is congruent modulo 19, we get

$$S' = \{11, 3, 14, 6, 17, 9, 1, 12, 4\}.$$

There are four members, namely 11, 12, 14 and 17, that exceed $\frac{19}{2}$. Hence, according to Gauss's Lemma,

$$(11/19) = (-1)^4 = 1,$$

and we can conclude that 11 is a quadratic residue of 19.

Suppose now that we replace the four members of S' that exceed $\frac{19}{2}$ by their negatives; that is, replace

$$11 \text{ by } -11 \equiv 8, \quad 12 \text{ by } -12 \equiv 7, \quad 14 \text{ by } -14 \equiv 5 \quad \text{and}$$
$$17 \text{ by } -17 \equiv 2 \pmod{19}.$$

This results in the new set

$$S'' = \{8, 3, 5, 6, 2, 9, 1, 7, 4\}.$$

Notice that S'' consists of the first nine natural numbers. This is typical of the general situation. The crux of the proof of Gauss's Lemma lies in showing that, for any odd prime p and coprime integer a, the set S'', constructed from multiples of a in this way, consists of the first $\frac{p-1}{2}$ natural numbers. It is a short step from there to completion of the proof.

Returning to our specific example, notice that the product of the elements in S'' is $9!$ while the product of the elements in S is $11^9 \times 9!$. Now the corresponding elements of the sets S and S'' are congruent modulo 19, with the exception of the four sign changes involved in getting from S' to S''. It follows that

$$11^9 \times 9! \equiv (-1)^4 \times 9! \pmod{19}$$

and so, cancelling the 9!, which is permissible since $\text{hcf}(9!, 19) = 1$,

$$11^9 \equiv 1 \pmod{19}.$$

Euler's Criterion now confirms that 11 is indeed a quadratic residue of 19.

The key to the calculation is the number of sign changes in getting from S' to S''. Had this turned out to be an odd integer then we would have had a minus sign remaining in the final congruence and, using Euler's Criterion, we would have concluded that the integer was a quadratic non-residue.

Our proof of Gauss's Lemma retraces the steps of this example, working, generally, with any odd prime p and coprime integer a.

Proof of Gauss's Lemma (Theorem 3.1) We are given that p is an odd prime, $a \not\equiv 0 \pmod{p}$ and

$$S = \left\{a, 2a, 3a, \ldots, \tfrac{p-1}{2}a\right\}.$$

First we observe that no two elements of S are congruent modulo p because, if $ra \equiv sa \pmod{p}$ for some integers r and s in the range $1 \leq r, s \leq \frac{p-1}{2}$, cancellation would give $r \equiv s \pmod{p}$, which can happen only if $r = s$. Moreover no element of S is congruent modulo p to 0, because $ra \equiv 0 \pmod{p}$, where $a \not\equiv 0 \pmod{p}$ requires $r \equiv 0 \pmod{p}$, which cannot occur.

Let S' be the set obtained by replacing each member of S by its least positive residue modulo p. Putting the elements of S' into increasing order, we can write

$$S' = \{b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_n\},$$

where $b_m < \frac{p}{2} < c_1$ and $m + n = \frac{p-1}{2}$. Now let

$$S'' = \{b_1, b_2, \ldots, b_m, p - c_1, p - c_2, \ldots, p - c_n\}.$$

Note that $\frac{p}{2}$ is not an integer and so neither b_m nor c_1 can be equal to $\frac{p}{2}$.

It is our contention that $S'' = \left\{1, 2, \ldots, \frac{p-1}{2}\right\}$. Certainly the integers listed in S'' are all positive, there are $\frac{p-1}{2}$ of them, and the largest is either b_m or $p - c_1$, each of which is less than $\frac{p}{2}$. Hence there are $\frac{p-1}{2}$ natural numbers each less than $\frac{p}{2}$, and so to justify our contention we need only show that they are distinct.

Note that this does not imply that $b_1 = 1$, $b_2 = 2$, etc. as the elements may be written in a different order.

As the elements of S' are distinct, the m elements $b_1, b_2, \ldots, b_m$ are distinct, as are the n elements $p - c_1, p - c_2, \ldots, p - c_n$. So suppose that $b_i = p - c_j$, for some i and j. Then

$$p = b_i + c_j \equiv ra + sa \pmod{p},$$

for some integers r and s in the range $1 \leq r, s \leq \frac{p-1}{2}$. But from

$$p \equiv (r + s)a \pmod{p} \text{ and } a \not\equiv 0 \pmod{p},$$

Euclid's Lemma leads to $r + s \equiv 0 \pmod{p}$, which is impossible since $2 \leq r + s \leq p - 1$.

This establishes that

$$S'' = \left\{1, 2, 3, \ldots, \tfrac{p-1}{2}\right\}.$$

We continue by multiplying the elements of S'' together:

$$\begin{aligned}\left(\tfrac{p-1}{2}\right)! &= b_1 b_2 \cdots b_m (p-c_1)(p-c_2)\cdots(p-c_n)\\ &\equiv b_1 b_2 \cdots b_m (-c_1)(-c_2)\cdots(-c_n) \pmod p\\ &\equiv (-1)^n b_1 b_2 \cdots b_m c_1 c_2 \cdots c_n \pmod p\\ &\equiv (-1)^n a \times 2a \times 3a \times \cdots \times \tfrac{p-1}{2}a \pmod p\\ &\equiv (-1)^n \times a^{(p-1)/2} \times \left(\tfrac{p-1}{2}\right)! \pmod p.\end{aligned}$$

This penultimate step follows from the fact that the elements of S' are congruent, in some order, to the elements of S.

Now $\mathrm{hcf}\left(p, \left(\frac{p-1}{2}\right)!\right) = 1$ and so the term $\left(\frac{p-1}{2}\right)!$ may be cancelled from the two sides of the congruence, leaving

$$1 \equiv (-1)^n a^{(p-1)/2} \pmod p.$$

Finally we appeal to Euler's Criterion in the form of property (d) of Theorem 2.2, that $a^{(p-1)/2} \equiv (a/p) \pmod p$, and then multiply both sides of the congruence by $(-1)^n$ to obtain

$$(a/p) = (-1)^n. \quad \blacksquare$$

Gauss's Lemma can be used in a very straightforward way to determine Legendre symbols: just write down the set S, convert it to the set S' and count how many members of S' lie in the appropriate range. Providing the prime modulus is not large, the arithmetic involved in doing this is not taxing.

Exercise 3.1

Use Gauss's Lemma to evaluate the following Legendre symbols.

(a) $(7/13)$

(b) $(11/23)$

(c) $(5/29)$

3.2 Applications of Gauss's Lemma

Property (b) of Theorem 2.2 establishes that $(1/p) = 1$, and so we need not consider $n = 1$ further.

In order to evaluate (n/p) for any natural number $n > 1$ and odd prime p that does not divide n, consider the prime decomposition

$$n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}.$$

Repeated application of the multiplicative property of the Legendre symbol, property (c) of Theorem 2.2, leads to

$$(n/p) = (p_1/p)^{k_1}(p_2/p)^{k_2}\cdots(p_r/p)^{k_r}.$$

Now in considering Legendre symbols we must not overlook negative integers. Note that

$$-n = (-1)p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$$

and using the multiplicative property again we have

$$(-n/p) = (-1/p)(p_1/p)^{k_1}(p_2/p)^{k_2}\cdots(p_r/p)^{k_r}.$$

So, to evaluate (a/p) for any integer a not divisible by the odd prime p, it suffices to be able to calculate $(-1/p)$, $(2/p)$ and (p_i/p) for any odd prime $p_i \neq p$. We have already dealt with $(-1/p)$ in property (e) of Theorem 2.2. Evaluation of the general (p_i/p) will be postponed until we have the Law of Quadratic Reciprocity at our disposal. However, Gauss's Lemma will resolve the $(2/p)$ problem for us.

As $-1 \equiv p-1 \pmod{p}$ we can do without $(-1/p)$, however, using it is often a more efficient way to proceed.

Proposition 3.2 *The quadratic character of 2*

If p is an odd prime then

$$(2/p) = \begin{cases} 1, & \text{if } p \equiv 1 \pmod 8 \text{ or } p \equiv 7 \pmod 8, \\ -1, & \text{if } p \equiv 3 \pmod 8 \text{ or } p \equiv 5 \pmod 8. \end{cases}$$

The result can be expressed in either of the alternative forms

$$(2/p) = \begin{cases} 1, & \text{if } p \equiv \pm 1 \pmod 8, \\ -1, & \text{if } p \equiv \pm 3 \pmod 8, \end{cases}$$

or $(2/p) = (-1)^{(p^2-1)/8}$.

Proof To bring Gauss's Lemma to bear, we observe that $(2/p) = (-1)^n$, where n is the number of integers in the set

$$S = \left\{2\times 1,\ 2\times 2,\ 2\times 3,\ \ldots,\ 2\times\tfrac{p-1}{2}\right\}$$
$$= \{2, 4, 6, \ldots, p-1\}$$

that have least positive residue modulo p exceeding $\frac{p}{2}$. The integers in S are, as they stand, least positive residues of p and so our task is to count how many of them exceed $\frac{p}{2}$.

Now $2k > \frac{p}{2}$ if, and only if, $k > \frac{p}{4}$. It follows that the first $\operatorname{int}\left(\frac{p}{4}\right)$ multiples of 2 do not exceed $\frac{p}{2}$, where $\operatorname{int}(x)$ denotes the integer part of x; that is, the largest integer that does not exceed x.

The remaining multiples do exceed $\frac{p}{2}$. So

$$n = \tfrac{p-1}{2} - \operatorname{int}\left(\tfrac{p}{4}\right).$$

There are four cases to consider, corresponding to the four possible congruence classes modulo 8 of the odd prime p.

(i) $p = 8k+1 \Longrightarrow n = 4k - \operatorname{int}\left(2k+\frac{1}{4}\right) = 4k - 2k = 2k.$

(ii) $p = 8k+3 \Longrightarrow n = 4k+1 - \operatorname{int}\left(2k+\frac{3}{4}\right) = 4k+1-2k = 2k+1.$

(iii) $p = 8k+5 \Longrightarrow n = 4k+2 - \operatorname{int}\left(2k+\frac{5}{4}\right) = 4k+2-(2k+1) = 2k+1.$

(iv) $p = 8k+7 \Longrightarrow n = 4k+3 - \operatorname{int}\left(2k+\frac{7}{4}\right) = 4k+3-(2k+1) = 2k+2.$

In cases (i) and (iv), n is even and so $(2/p) = 1$. In cases (ii) and (iii), n is odd and so $(2/p) = -1$. ■

Exercise 3.2

(a) Write down the values of the following Legendre symbols.

(i) $(2/29)$ (ii) $(2/31)$ (iii) $(2/43)$

(b) Adapt the proof of Proposition 3.2 to show that

$$(-2/p) = \begin{cases} 1, & \text{if } p \equiv 1 \pmod 8 \text{ or } p \equiv 3 \pmod 8, \\ -1, & \text{if } p \equiv 5 \pmod 8 \text{ or } p \equiv 7 \pmod 8. \end{cases}$$

(c) Check the result of part (b) by confirming that

$$(-2/p) = (-1/p)(2/p)$$

holds.

The result of Proposition 3.2 provides us with a neat proof of Proposition 2.4 of Chapter 9. The result in question concerned the Mersenne numbers $M_n = 2^n - 1$.

This result was stated without proof in Chapter 9.

Proposition 3.3 *(Proposition 2.4 of Chapter 9)*

If p and $2p + 1$ are both primes, where $p \equiv 3 \pmod 4$, then $2p + 1$ divides M_p.

Proof Let q be the prime $2p + 1$. From $p = 4k + 3$ we have $2p + 1 = 8k + 7$ and so $q \equiv 7 \pmod 8$. Now Proposition 3.2 gives $(2/q) = 1$, so that 2 is a quadratic residue of q. Euler's Criterion gives $2^{(q-1)/2} \equiv (2/q) \pmod q$, and putting the two together gives $2^{(q-1)/2} \equiv 1 \pmod q$. That is, q divides $2^{(q-1)/2} - 1$, which, when q is replaced by $2p + 1$, says that $2p + 1$ divides M_p, as claimed. ■

As we saw in Chapter 9, this result shows that M_{11}, M_{23}, M_{83} and numerous other Mersenne numbers are composite. It remains an unsolved problem whether the list we have begun here extends infinitely; that is, it is not known whether there are infinitely many primes p of the form $4k + 3$ for which $2p + 1$ is also prime.

3.3 More cases of Dirichlet's Theorem

In Book A, Chapter 2 we mentioned, without proof, Dirichlet's Theorem, which asserts that, for coprime natural numbers a and b, there are infinitely many primes of the form $ak + b$. We proved some special instances of this result, including the infinitude of primes of the form $4k + 3$. In a curious way our results on quadratic residues will lead to proofs of many more particular instances of Dirichlet's result. As Chapter 2 posed the question of whether or not there are infinitely many primes of the form $4k + 1$, this is the obvious starting point for us here.

Proposition 3.4

There are infinitely many primes of the form $4k + 1$.

Proof Aiming for a contradiction, suppose that there are only a finite number of primes of the form $4k + 1$, say $p_1, p_2, \ldots, p_n$.

Let $N = (2p_1p_2 \cdots p_n)^2 + 1$. As N is odd it must be divisible by some odd prime, say p. Then $(2p_1p_2 \cdots p_n)^2 \equiv -1 \pmod{p}$. As the term on the left of this congruence is a square we can infer that $(-1/p) = 1$ and then, by virtue of property (e) of Theorem 2.2, that p is of the form $4k + 1$. The assumption that $p_1, p_2, \ldots, p_n$ are the only primes of this form forces $p = p_i$, for some i. But this implies that p divides $(2p_1p_2 \cdots p_n)^2$, and coupling this fact with p divides N we conclude that p divides $N - (2p_1p_2 \cdots p_n)^2$. That is, p divides 1, and we have the required contradiction.

Hence there are infinitely many primes of the form $4k + 1$. ■

Although we presented the proof of Proposition 3.4 as one by contradiction, the essence of the argument was to show that from any list of primes of the form $4k + 1$ we can always discover another prime of the same form. From the list $p_1, p_2, \ldots, p_n$ we constructed $(2p_1p_2 \cdots p_n)^2 + 1$, which has to be divisible by a prime of form $4k + 1$ that is not one of the original list. Many special instances of Dirichlet's result can be proved in more or less this way, the key to the proof being in finding the 'new' number to construct. In our next example we use a similar method, but with an extra step like the one used previously in our corresponding proof that there are infinitely many primes of the form $4k + 3$.

Proposition 3.5

There are infinitely many primes of the form $8k - 1$.

Integers of the form $8k - 1$ are equally well expressed as integers of the form $8k + 7$.

Proof Suppose that there are only finitely many primes of the form $8k-1$. If there are only finitely many then there is a largest such prime, say P. Consider the integer $N = 2(P!)^2 - 1$. As $P!$ is even, $2(P!)^2$ is divisible by 8 and so $N \equiv -1 \pmod 8$.

Being odd, N has an odd prime factor. Let p be such a factor. Then $2(P!)^2 \equiv 1 \pmod p$. So

$$(2/p)((P!)^2/p) = (1/p),$$

and as $(P!)^2$ and 1 are both squares we deduce that $(2/p) = 1$. Bringing Proposition 3.2 into play, p is congruent modulo 8 to either 1 or -1.

Now any product of integers congruent modulo 8 to 1 is itself congruent modulo 8 to 1. As N is not of this form, at least one of its prime factors, say q, must be congruent modulo 8 to -1. The assumption that the largest prime of this form is P tells us that $q \leq P$. But then q divides $2(P!)^2$ and as q divides $2(P!)^2 - 1$, we must have q divides 1, which is a contradiction.

Hence there are infinitely many primes of the form $8k-1$. ■

Exercise 3.3

Show that there are infinitely many primes of the form $8k+3$ by the following method. Suppose that there are only a finite number of primes of the form $8k+3$, namely $p_1, p_2, \ldots, p_n$. Consider the integer $N = (p_1 p_2 \cdots p_n)^2 + 2$. For any prime factor p of N, show that $(-2/p) = 1$ and deduce that N must have a prime factor of the form $8k+3$. Reach a contradiction by noting that this prime would have to be one of the original list $p_1, p_2, \ldots, p_n$.

Exercise 3.4

(a) Use Gauss's Lemma to determine the following Legendre symbols.

(i) $(-3/13)$ (ii) $(5/17)$

(b) Show that 2 is a quadratic residue of:

(i) every prime of the form $4^n + 1$, for $n > 1$

(ii) every Mersenne prime $M_p = 2^p - 1$, except $M_2 = 3$

(iii) every Fermat prime $F_n = 2^{2^n} + 1$, for $n > 1$.

For each of these primes, decide whether or not -2 is a quadratic residue.

(c) Show that there are infinitely many primes of the form $8k+5$ in the following way. Suppose that $p_1, p_2, \ldots, p_n$ are the only primes of this form, and reach a contradiction by considering the integer

$$N = 4(p_1 p_2 \cdots p_n)^2 + 1.$$

4 The Law of Quadratic Reciprocity

The Law of Quadratic Reciprocity, described by Gauss as the 'gem of higher arithmetic', is the third and final result of this chapter.

4.1 Understanding the Law

This section is concerned with the Law of Quadratic Reciprocity, which, for ease of reference, we will abbreviate to LQR. Amongst its numerous applications, LQR provides us with a new method for simplifying Legendre symbols. However, unlike the preceding methods, LQR is simple to employ even when the integers in question are large.

Before we confront the difficult task of proving LQR, we begin by illustrating the implications of the result through some simple worked exercises.

Theorem 4.1 *The Law of Quadratic Reciprocity*

If p and q are distinct odd primes then

$$(p/q)(q/p) = (-1)^{(p-1)(q-1)/4}.$$

This bland statement probably does not mean much to you yet. Let us hold fire for a moment and express the statement of LQR in a slightly different way to clarify just what it entails. Look at the expression $(-1)^{(p-1)(q-1)/4}$ and suppose that $p \equiv 1 \pmod 4$. In this case $p-1$ is divisible by 4 and, since $q-1$ is also even, $\dfrac{(p-1)(q-1)}{4}$ is an even integer. Therefore $(-1)^{(p-1)(q-1)/4} = 1$. Similarly if $q \equiv 1 \pmod 4$ then $(-1)^{(p-1)(q-1)/4} = 1$. On the other hand, if $p \equiv q \equiv 3 \pmod 4$ then $p-1$ and $q-1$ are each equal to twice an odd integer and so $\dfrac{(p-1)(q-1)}{4}$ is odd. In this case $(-1)^{(p-1)(q-1)/4} = -1$. The statement of LQR can now be given in the alternative form

$$(p/q)(q/p) = \begin{cases} 1, & \text{if either } p \equiv 1 \pmod 4 \quad \text{or} \quad q \equiv 1 \pmod 4, \\ -1, & \text{if } p \equiv q \equiv 3 \pmod 4. \end{cases}$$

But we can take this a step further. As (p/q) and (q/p) are each equal to either 1 or -1, their product $(p/q)(q/p)$ will be equal to 1 precisely when $(p/q) = (q/p)$ and will be equal to -1 precisely when $(p/q) = -(q/p)$. We therefore have the following way of stating LQR, which will prove to be a more useful way of looking at it in many practical situations.

Theorem 4.2 *An alternative formulation of LQR*

If p and q are distinct odd primes then

$$(p/q) = \begin{cases} (q/p), & \text{if either } p \equiv 1 \pmod 4 \quad \text{or} \quad q \equiv 1 \pmod 4, \\ -(q/p), & \text{if } p \equiv q \equiv 3 \pmod 4. \end{cases}$$

LQR provides the missing link in the processes required to evaluate any Legendre symbol (a/p), where a is an integer not divisible by the odd prime p. We have seen that it is sufficient to be able to evaluate $(-1/p), (2/p)$ and (q/p), for any odd prime $q \neq p$. We have found simple formulas for $(-1/p)$ and $(2/p)$. LQR now provides the means of determining (q/p). We illustrate the method using a worked exercise.

Worked Exercise 4.3

Is 31 a quadratic residue of 73?

Solution

We need to evaluate the Legendre symbol $(31/73)$.

$$\begin{aligned} (31/73) &= (73/31), && \text{by LQR, noting that } 73 \equiv 1 \pmod 4, \\ &= (11/31), && \text{since } 73 \equiv 11 \pmod{31}, \\ &= (-1)(31/11), && \text{by LQR,} \\ & && \text{noting that } 11 \equiv 31 \equiv 3 \pmod 4, \\ &= (-1)(9/11), && \text{since } 31 \equiv 9 \pmod{11}, \\ &= -1, && \text{since 9 is a square.} \end{aligned}$$

So 31 is not a quadratic residue of 73.

The underlying process illustrated in this worked exercise is that the evaluation of any Legendre symbol is achieved by replacing it with a product of one or more simpler Legendre symbols – simpler in the sense that the integers involved in the symbol are smaller. LQR is needed to simplify the Legendre symbol (p/q), where $p < q$, by replacing it with either (q/p) or $(-1)(q/p)$, whichever is the case. As $p < q$, the symbol (q/p) can then be simplified by, for example, replacing q with its least positive residue modulo p (which is allowed by property (a) of Theorem 2.2). As the integers are repeatedly reduced, a stage will be reached when the value of the Legendre symbol is known.

The next example is similar, but involves a larger prime, 229. We have illustrated the breakdown into 'smaller' Legendre symbols schematically.

Worked Exercise 4.4

Determine the Legendre symbol $(-134/229)$.

Solution

$$
\begin{array}{ccccc}
 & & (-134/229) & & \\
 & & \downarrow{\scriptstyle 2.2(c)} & & \\
(-1/229) & \times & (2/229) & \times & (67/229) \\
\downarrow{\scriptstyle 2.2(e)} & & \downarrow{\scriptstyle 3.2} & & \downarrow{\scriptstyle \text{LQR}} \\
1 & \times & (-1) & \times & (229/67) \\
 & & & & \downarrow{\scriptstyle 2.2(a)} \\
 & & & & (28/67) \\
 & & & & \downarrow{\scriptstyle 2.2(c)} \\
 & & & (4/67) & \times \quad (7/67) \\
 & & & \downarrow{\scriptstyle 2.2(b)} & \qquad \downarrow{\scriptstyle \text{LQR}} \\
 & & & 1 & \times \quad (-1) \times (67/7) \\
 & & & & \qquad \downarrow{\scriptstyle 2.2(a)} \\
 & & & & \qquad (4/7) \\
 & & & & \qquad \downarrow{\scriptstyle 2.2(b)} \\
 & & & & \qquad 1
\end{array}
$$

So $(-134/229) = 1 \times (-1) \times 1 \times (-1) \times 1 = 1$.

Exercise 4.1

(a) Determine the following Legendre symbols.

(i) $(-219/383)$ (ii) $(461/773)$

The integers 383, 461 and 773 are primes.

(b) Explain each of the steps in the following evaluation of the Legendre symbol $(164/197)$.

$$\begin{aligned}(164/197) &= (4/197)(41/197)\\ &= (41/197)\\ &= (197/41)\\ &= (74/41)\\ &= (2/41)(37/41)\\ &= (37/41)\\ &= (41/37)\\ &= (4/37)\\ &= 1\end{aligned}$$

(c) The following evaluation of $(60/79)$ contains a number of errors involving misuse of LQR and properties of the Legendre symbol. Identify the errors.

$$\begin{aligned}(60/79) &= (-19/79), \quad \text{since } 60 \equiv -19 \,(\text{mod } 79),\\ &= (-1)(19/79), \quad \text{since } -19 = (-1) \times 19,\\ &= (-1)(79/19), \quad \text{by LQR},\\ &= (-1)(60/19), \quad \text{since } 79 \equiv 60 \,(\text{mod } 19),\\ &= (-1)(4/19)(15/19), \quad \text{since } 60 = 4 \times 15,\\ &= (-1)(15/19), \quad \text{since 4 is a square},\\ &= (-1)(-1)(19/15), \quad \text{by LQR,}\\ &\qquad\qquad \text{noting that } 15 \equiv 19 \equiv 3 \,(\text{mod } 4),\\ &= (4/15), \quad \text{since } 19 \equiv 4 \,(\text{mod } 15),\\ &= 1, \quad \text{since 4 is a square.}\end{aligned}$$

4.2 Proof of LQR

Since the first proof, given by Gauss, over 100 further proofs have been published. None of them is short and snappy! The one that we have opted to give here was discovered by one of Gauss's students, Ferdinand Eisenstein (1823–1852); like many of the published proofs, it builds on Gauss's Lemma. The proof is really in two distinct parts, the first of which establishes the link between LQR and Gauss's Lemma. To make the proof easier to understand, we establish the first part as a separate result.

Lemma 4.5

Let p be an odd prime and a an odd integer not divisible by p. Then

$$(a/p) = (-1)^{\alpha(a,p)},$$

where

$$\alpha(a,p) = \sum_{k=1}^{(p-1)/2} \text{int}\left(\frac{ka}{p}\right).$$

α is the Greek letter 'alpha'.

To clarify the statement of the lemma, take the case $a = 5$ and $p = 13$. Then $\alpha(5,13)$ comprises the sum of $\frac{13-1}{2} = 6$ terms:

$$\begin{aligned}\alpha(5,13) &= \text{int}\left(\tfrac{5}{13}\right) + \text{int}\left(\tfrac{10}{13}\right) + \text{int}\left(\tfrac{15}{13}\right) + \text{int}\left(\tfrac{20}{13}\right) + \text{int}\left(\tfrac{25}{13}\right) + \text{int}\left(\tfrac{30}{13}\right)\\ &= 0+0+1+1+1+2 = 5,\end{aligned}$$

and so, according to the lemma, $(5/13) = (-1)^5 = -1$.

Proof Exactly as in our proof of Gauss's Lemma, we consider the set of integers

$$S = \left\{a, 2a, 3a, \ldots, \tfrac{p-1}{2}a\right\}$$

and, replacing each integer with its least positive residue modulo p and listing the integers in ascending order, we reach

$$S' = \{b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_n\},$$

where $b_m < \frac{p}{2} < c_1$.

Now, according to the Division Algorithm, if we divide any $ka \in S$ by p we get

$$ka = p \times \text{int}\left(\frac{ka}{p}\right) + \text{remainder},$$

where the remainder is one of the integers in S'. As in the proof of Gauss's Lemma, the remainders that occur are distinct members of S'. Allowing k to run from 1 to $\frac{p-1}{2}$ and adding up the resulting equations gives

$$\sum_{k=1}^{(p-1)/2} ka = \sum_{k=1}^{(p-1)/2} p \times \text{int}\left(\frac{ka}{p}\right) + \sum_{i=1}^{m} b_i + \sum_{j=1}^{n} c_j. \qquad (*)$$

In the proof of Gauss's Lemma we went on to show that the set

$$\begin{aligned}S'' &= \{b_1, b_2, \ldots, b_m, p-c_1, p-c_2, \ldots, p-c_n\}\\ &= \left\{1, 2, 3, \ldots, \tfrac{p-1}{2}\right\}.\end{aligned}$$

Adding the integers in the set S'' gives

$$\sum_{k=1}^{(p-1)/2} k = \sum_{i=1}^{m} b_i + \sum_{j=1}^{n} (p - c_j) = \sum_{i=1}^{m} b_i + pn - \sum_{j=1}^{n} c_j.$$

Subtracting this equation from $(*)$ gives

$$\sum_{k=1}^{(p-1)/2} ka - \sum_{k=1}^{(p-1)/2} k = \sum_{k=1}^{(p-1)/2} p \times \operatorname{int}\left(\frac{ka}{p}\right) + 2 \times \sum_{j=1}^{n} c_j - pn.$$

That is,

$$(a-1) \sum_{k=1}^{(p-1)/2} k = p \times \alpha(a,p) + 2 \times \sum_{j=1}^{n} c_j - pn.$$

Finally we write this equation modulo 2, noting that each of the terms still involving summation is even (since a is odd), to get

$$p\alpha(a,p) - pn \equiv 0 \,(\mathrm{mod}\ 2).$$

Since p is odd this amounts to $\alpha(a,p) \equiv n \,(\mathrm{mod}\ 2)$. From the conclusion of Gauss's Lemma that $(a/p) = (-1)^n$ we have reached $(a/p) = (-1)^{\alpha(a,p)}$, as claimed. ■

With this result at our disposal, it is just a short step to completion of the proof of LQR.

Proof of LQR (Theorem 4.1) For the given odd primes p and q, consider the rectangle in the xy-plane that has vertices at $(0,0)$, $\left(\frac{p}{2},0\right)$, $\left(\frac{p}{2},\frac{q}{2}\right)$ and $\left(0,\frac{q}{2}\right)$. The number of lattice points *inside* this rectangle is $\frac{p-1}{2} \times \frac{q-1}{2}$. Consider the diagonal from (0,0) to $\left(\frac{p}{2},\frac{q}{2}\right)$, which has equation $y = \frac{q}{p}x$. Inside the rectangle there are no lattice points on this diagonal, because if (a,b) is a lattice point on this diagonal then $b = \frac{q}{p}a$. But as p and q are distinct primes, this implies that p divides a and q divides b, which forces (a,b) outside the rectangle.

Lattice points are points (a,b), where each of a and b is an integer.

Denote by A and B the triangular regions inside the rectangle that lie respectively above and below the diagonal (Figure 4.1).

For integer k with $0 < k < \frac{p}{2}$, the number of lattice points in B lying directly above the point $(k,0)$ on the x-axis is $\operatorname{int}\left(\frac{q}{p}k\right)$. Hence the total number of lattice points in B is

$$\sum_{k=1}^{(p-1)/2} \operatorname{int}\left(\frac{q}{p}k\right),$$

which is $\alpha(q,p)$. In the same way, by counting the lattice points to the right of $(0,k)$, the total number of lattice points in A is $\alpha(p,q)$. Now the number of lattice points in A added to the number of lattice points in B is the total number of lattice points in the rectangle, which is

The argument is symmetric with the roles of p and q interchanged.

$$\frac{p-1}{2} \times \frac{q-1}{2} = \frac{(p-1)(q-1)}{4}.$$

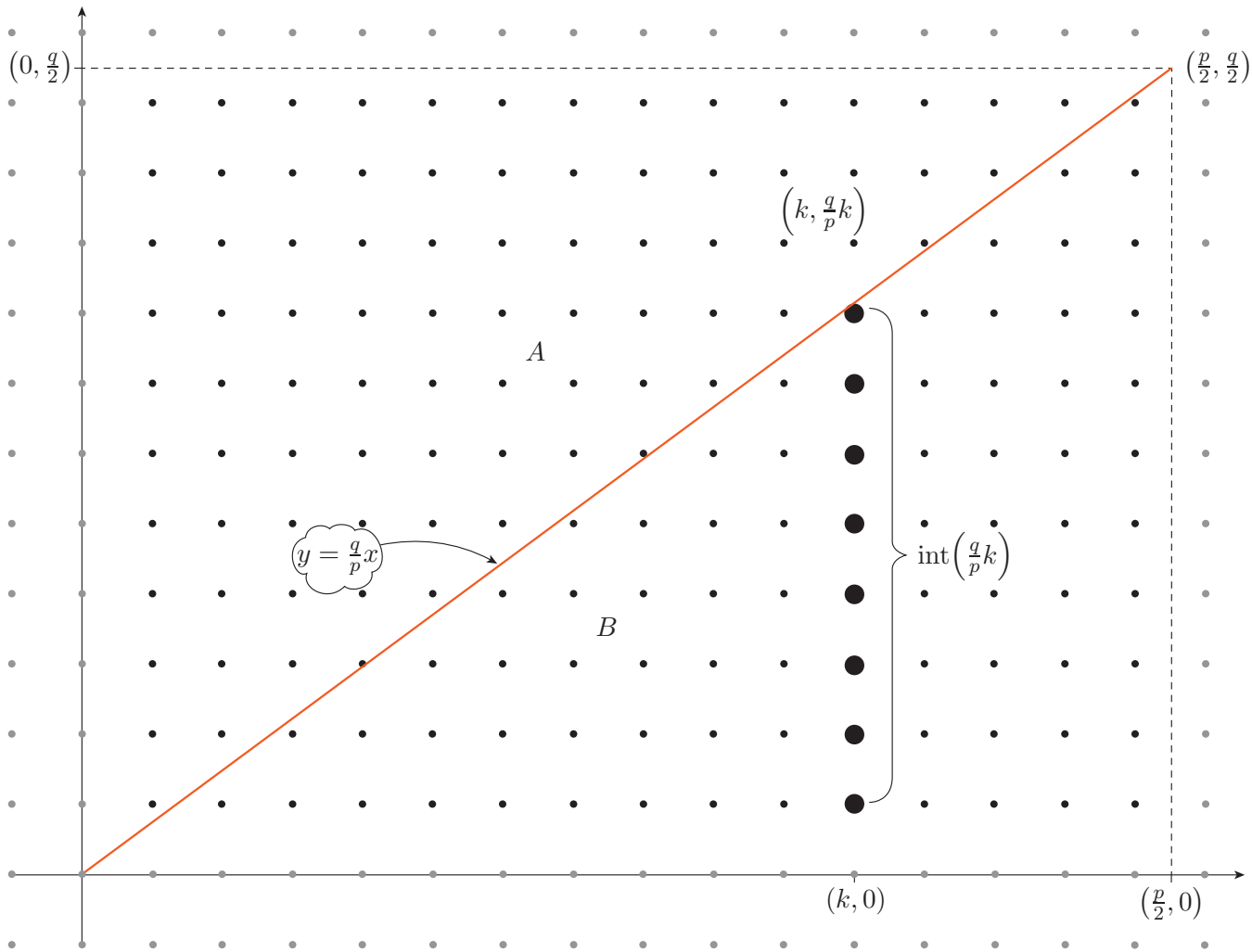


Figure 4.1 Counting the lattice points above and below the diagonal from $(0, 0)$ to $(\frac{p}{2}, \frac{q}{2})$

Hence, from the two ways of counting the number of lattice points,

$$\alpha(p, q) + \alpha(q, p) = \frac{(p-1)(q-1)}{4}.$$

The equality of the two counts uses the fact that there are no points on the diagonal.

So, in terms of Legendre symbols, by Lemma 4.5,

$$\begin{aligned}(p/q)(q/p) &= (-1)^{\alpha(p,q)} \times (-1)^{\alpha(q,p)} \\ &= (-1)^{\alpha(p,q)+\alpha(q,p)} \\ &= (-1)^{(p-1)(q-1)/4},\end{aligned}$$

which is LQR. ■

4.3 Applications of LQR

To consolidate your appreciation of LQR, let us round off the section with some more examples using it. We started this chapter with a question on the solvability or otherwise of a given quadratic congruence. Let us return to that theme.

Worked Exercise 4.6

139 is prime.

Does the congruence $3x^2 - 8x - 5 \equiv 0 \pmod{139}$ have any solutions?

Solution

The discriminant of the quadratic is

$$(-8)^2 - 4 \times 3 \times (-5) = 124,$$

and so the question boils down to the determination of the Legendre symbol $(124/139)$. In what follows we use, in addition to LQR, the various Legendre symbol properties of Theorem 2.2. We have given an explanation of each step rather than a reference to the property being used.

$$\begin{aligned}
(124/139) &= (4/139)(31/139), \quad \text{using } 124 = 4 \times 31,\\
&= (1)(31/139), \quad \text{since } 4 = 2^2,\\
&= (-1)(139/31), \quad \text{by LQR, since } 31 \equiv 139 \equiv 3 \pmod 4,\\
&= (-1)(15/31), \quad \text{since } 139 \equiv 15 \pmod{31},\\
&= (-1)(3/31)(5/31), \quad \text{using } 15 = 3 \times 5,\\
&= (-1)(-1)(31/3)(5/31), \quad \text{by LQR,}\\
&\qquad\qquad \text{since } 31 \equiv 3 \pmod 4,\\
&= (31/3)(31/5), \quad \text{by LQR, since } 5 \equiv 1 \pmod 4,\\
&= (1/3)(31/5), \quad \text{since } 31 \equiv 1 \pmod 3,\\
&= (1/3)(1/5), \quad \text{since } 31 \equiv 1 \pmod 5,\\
&= 1.
\end{aligned}$$

It follows that the congruence does have solutions.

Exercise 4.2

For each of the following congruences, decide whether or not it has any solutions.

(a) $3x^2 + 6x + 5 \equiv 0 \pmod{89}$

(b) $2x^2 + 5x - 9 \equiv 0 \pmod{11 \times 101}$

Proposition 3.2 established a simple formula for determining the quadratic character of 2. Is there a similar result for determining the quadratic character of 3? In terms of the Legendre symbol, we are asking: which primes $p \geq 5$ satisfy $(3/p) = 1$? LQR gives us a start in tackling this one. Since $3 \equiv 3 \pmod 4$, the result of applying LQR to $(3/p)$ depends on the value of p modulo 4:

$$(3/p) = \begin{cases} (p/3), & \text{if } p \equiv 1 \pmod 4, \\ (-1)(p/3), & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

Now the least positive residue of p modulo 3 is either 1 (which is a square) or 2 (which is not), and so

$$(p/3) = \begin{cases} 1, & \text{if } p \equiv 1 \pmod 3, \\ -1, & \text{if } p \equiv 2 \pmod 3. \end{cases}$$

The two parts need to be combined to obtain the value of $(3/p)$, and for that we need to consider values of p to both modulus 3 and modulus 4; so we look at cases modulo 12. For example, if $p \equiv 7 \pmod{12}$ then we have $p \equiv 3 \pmod 4$ and $p \equiv 1 \pmod 3$. Therefore

$$(3/p) = (-1)(p/3) \quad \text{and} \quad (p/3) = 1,$$

which combine to give $(3/p) = -1$.

There are, in all, four possible values modulo 12 that a prime that is greater than 3 can take, namely $p \equiv 1, 5, 7$ or $11 \pmod{12}$. The value of $(3/p)$ in each case is as determined in Table 4.1.

Table 4.1 Value of $(3/p)$ in the four possible cases

	$p \pmod 3$	$(p/3)$	$p \pmod 4$	$(3/p)$
$p \equiv 1 \pmod{12}$	1	1	1	1
$p \equiv 5 \pmod{12}$	2	-1	1	-1
$p \equiv 7 \pmod{12}$	1	1	3	-1
$p \equiv 11 \pmod{12}$	2	-1	3	1

We can summarise these findings as follows.

Proposition 4.7 *The quadratic character of 3*

If $p > 3$ is prime then

$$(3/p) = \begin{cases} 1, & \text{if } p \equiv \pm 1 \pmod{12}, \\ -1, & \text{if } p \equiv \pm 5 \pmod{12}. \end{cases}$$

Exercise 4.3

(a) For which primes $p > 3$ is $(6/p) = 1$?

(b) Prove that, for prime $p > 3$, $(-3/p) = 1$ if, and only if, $p \equiv 1 \pmod 6$. Use this to prove that there are infinitely many primes of the form $6k + 1$ by considering the integer $(2p_1p_2 \cdots p_n)^2 + 3$, where p_1, p_2, ..., p_n are assumed to be the only primes of the form $6k + 1$.

We have achieved the main goal of this chapter. Given any quadratic congruence with prime modulus, we can decide whether or not it has solutions. Moreover, part (b) of Exercise 4.2 illustrates a fruitful line of approach for similar problems where the modulus is a product of distinct primes. As far as general composite moduli are concerned, there is one idea of particular interest in our present situation that merits mention. The Legendre symbol can be generalised in a simple way.

Definition 4.8 *The Jacobi symbol*

Let m and n be integers, with n odd, $n \geq 3$ and $\text{hcf}(m, n) = 1$. If we write $n = p_1p_2 \cdots p_r$ as a product of (not necessarily distinct) primes, then the **Jacobi symbol** (m/n) is defined by

$$(m/n) = (m/p_1)(m/p_2) \cdots (m/p_r),$$

where the symbols on the right-hand side are Legendre symbols.

So, for example,

$$(14/275) = (14/5)(14/5)(14/11) = (4/5)(4/5)(3/11) = 1 \times 1 \times 1 = 1.$$

By Proposition 4.7, $(3/11) = 1$.

From our knowledge of Legendre symbols there is little remaining challenge in evaluating Jacobi symbols. But what does the Jacobi symbol represent? It would be nice if, for odd coprime integers m and n, we had $(m/n) = 1$ if, and only if, m is a non-zero square modulo n. Unfortunately this is not true. For example, 5 is not a square modulo 9, but

$$(5/9) = (5/3)(5/3) = (2/3)(2/3) = (-1)(-1) = 1.$$

We talk about 'squares' rather than quadratic residues because the latter are defined only for prime modulus.

But all is not lost, because it is true that if $(m/n) = -1$ then m is not a square modulo n. To see this, suppose that $(m/n) = -1$. Then in the expression of (m/n) as a product of Legendre symbols there is (at least) one prime factor p of n for which $(m/p) = -1$. That is, m is not a square modulo p. It follows that m cannot be a square modulo n.

If $x^2 \equiv m \pmod n$ then $x^2 \equiv m \pmod d$ for any factor d of n.

The Jacobi symbol (m/n) is a generalisation of the Legendre symbol in that the two notions coincide when n is prime. In fact, almost all the properties we have seen for the Legendre symbol do generalise. In particular, the following all hold true; we state them without proof.

Proposition 4.9

If m and n are odd integers both exceeding 1 with $\text{hcf}(m, n) = 1$, then:

(a) $(-1/n) = (-1)^{(n-1)/2}$, Euler's Criterion

(b) $(2/n) = (-1)^{(n^2-1)/8}$, Proposition 3.2

(c) $(m/n)(n/m) = (-1)^{(m-1)(n-1)/4}$, LQR.

Exercise 4.4

(a) Evaluate the following Legendre symbols.

(i) $(35/71)$ (ii) $(139/149)$ (iii) $(1234/4567)$

The integers 139, 149 and 4567 are all primes.

(b) Which of the following congruences have solutions?

(i) $3x^2 + x - 7 \equiv 0 \pmod{17}$

(ii) $2x^2 + x + 1 \equiv 0 \pmod{29}$

(iii) $2x^2 + 3x - 7 \equiv 0 \pmod{101}$

(c) (i) If p and $q = 10p + 3$ are odd primes, show that $(p/q) = (3/p)$.

(ii) If p and $q = 10p + 1$ are odd primes, show that $(p/q) = (-1/p)$.

(d) Show that $4^n \equiv 4 \pmod{12}$ for all integers $n \geq 1$. Use this result and Proposition 4.7 to show that 3 is a quadratic non-residue of each of the primes given in Exercise 3.4(b) parts (i), (ii) and (iii).

(e) For which odd primes $p \neq 7$ does $(7/p) = 1$?

(f) If p and q are distinct odd primes prove that $(-p/q) = (-q/p)$ if, and only if, $p \equiv q \equiv 1 \pmod{4}$.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should:

(a) Be able to solve quadratic congruences modulo a prime by completing the square, and understand the importance of the discriminant in determining whether a quadratic congruence has solutions.

(b) Understand the definition of quadratic residue and quadratic non-residue of a prime, and be able to find them for small primes.

(c) Understand identities involving quadratic residues, including Euler's Criterion and its uses.

(d) Recall the definition and basic properties of the Legendre symbol, and be able to use these properties to evaluate a variety of Legendre symbols.

(e) Know how to use the statement of Gauss's Lemma, and understand its proof and applications.

(f) Recognise how to prove certain special cases of Dirichlet's Theorem using Legendre symbols.

(g) Understand the Law of Quadratic Reciprocity, and be able to use it to find Legendre symbols.

(h) Recall the approach and key steps to proving the Law of Quadratic Reciprocity.

(i) Understand the applications of the Law of Quadratic Reciprocity to solving quadratic congruences and establishing the quadratic character of the number 3.

(j) Show awareness that results for the Legendre symbol can be extended to the Jacobi symbol.

Solutions to exercises

Solution to Exercise 1.1

(a) Multiplying through by 20 gives

$$100x^2 + 140x + 20 \equiv 0 \,(\text{mod } 13),$$

and completing the square gives

$$(10x + 7)^2 - 29 \equiv 0 \,(\text{mod } 13).$$

That is,

$$(10x + 7)^2 \equiv 29 \equiv 16 \,(\text{mod } 13).$$

As $16 = (\pm 4)^2$, the required solutions are obtained by solving $10x + 7 \equiv \pm 4 \,(\text{mod } 13)$. These are

$$x \equiv 1 \,(\text{mod } 13) \quad \text{and} \quad x \equiv 8 \,(\text{mod } 13).$$

(b) Multiply through by 4,

$$4x^2 + 4x + 4 \equiv 0 \,(\text{mod } 5),$$

and complete the square,

$$(2x + 1)^2 + 3 \equiv 0 \,(\text{mod } 5).$$

Add 2 to both sides:

$$(2x + 1)^2 \equiv 2 \,(\text{mod } 5).$$

But the squares modulo 5 are 0, 1 and 4, and so the congruence

$$y^2 \equiv 2 \,(\text{mod } 5)$$

has no solutions. Therefore the original congruence in x has no solutions.

(c) Multiplying through by 4 and completing the square gives

$$(2x + 3)^2 + 7 \equiv 0 \,(\text{mod } 7).$$

That is,

$$(2x + 3)^2 \equiv 0 \,(\text{mod } 7).$$

So this has a unique solution, which satisfies

$$2x + 3 \equiv 0 \,(\text{mod } 7),$$

namely $x \equiv 2 \,(\text{mod } 7)$.

Solution to Exercise 1.2

Working modulo 13,

$$1^2 \equiv 12^2 \equiv 1, \quad 2^2 \equiv 11^2 \equiv 4, \quad 3^2 \equiv 10^2 \equiv 9,$$
$$4^2 \equiv 9^2 \equiv 3, \quad 5^2 \equiv 8^2 \equiv 12, \quad 6^2 \equiv 7^2 \equiv 10.$$

The quadratic residues of 13 are therefore 1, 3, 4, 9, 10 and 12, and the non-residues are 2, 5, 6, 7, 8 and 11.

The congruence $x^2 + x - c \equiv 0 \pmod{13}$ has solution(s) when its discriminant, $1 + 4c$, is congruent to either 0 or a quadratic residue of 13.

$$\begin{aligned}
1 + 4c &\equiv 0 \pmod{13} &&\iff c \equiv 3 \pmod{13} \\
1 + 4c &\equiv 1 \pmod{13} &&\iff c \equiv 0 \pmod{13} \\
1 + 4c &\equiv 3 \pmod{13} &&\iff c \equiv 7 \pmod{13} \\
1 + 4c &\equiv 4 \pmod{13} &&\iff c \equiv 4 \pmod{13} \\
1 + 4c &\equiv 9 \pmod{13} &&\iff c \equiv 2 \pmod{13} \\
1 + 4c &\equiv 10 \pmod{13} &&\iff c \equiv 12 \pmod{13} \\
1 + 4c &\equiv 12 \pmod{13} &&\iff c \equiv 6 \pmod{13}
\end{aligned}$$

So the congruence has solutions when $c \equiv 0, 2, 3, 4, 6, 7$ or $12 \pmod{13}$.

Solution to Exercise 1.3

According to Euler's Criterion, 5 is a quadratic residue of prime p if, and only if, $5^{(p-1)/2} \equiv 1 \pmod{p}$.

$$5^{(11-1)/2} = 5^5 = 25 \times 25 \times 5 \equiv 3 \times 3 \times 5 = 45 \equiv 1 \pmod{11}$$
$$5^{(17-1)/2} = 5^8 = (5^2)^4 \equiv 8^4 = 64^2 \equiv (-4)^2 \equiv -1 \pmod{17}$$
$$5^{(19-1)/2} = 5^9 = 5 \times (5^2)^4 \equiv 5 \times 6^4 \equiv 5 \times (-2)^2 \equiv 1 \pmod{19}$$

So 5 is a quadratic residue of 11 and 19 but is a quadratic non-residue of 17.

Solution to Exercise 1.4

(a) Since, being a square, x^2 is a quadratic residue of p, Table 1.1 informs us that a and b must be of the same quadratic character. That is, a is a quadratic residue of p if, and only if, b is a quadratic residue of p.

(b) If -1 is a quadratic residue of p then, according to Euler's Criterion, $(-1)^{(p-1)/2} = 1$ and so $\frac{p-1}{2}$ is an even integer. That is, $p \equiv 1 \pmod{4}$.

(c) Since $p - a \equiv -a \equiv (-1)a \pmod{p}$, a and $p - a$ will be of the same quadratic character when -1 is a quadratic residue of p. Hence $p - a$ will also be a quadratic residue precisely when $p \equiv 1 \pmod{4}$.

Solution to Exercise 1.5

(a) (i) Multiplying through by 8 and completing the square, the congruence becomes

$$(4x+4)^2 - 8 \equiv 0 \,(\text{mod } 7), \text{ that is, } (4x+4)^2 \equiv 1 \,(\text{mod } 7).$$

This has solutions satisfying $4x + 4 \equiv \pm 1 \,(\text{mod } 7)$. That is, $x \equiv 1 \,(\text{mod } 7)$ and $x \equiv 4 \,(\text{mod } 7)$.

(ii) Writing the congruence as $3x^2 + 12x - 3 \equiv 0 \,(\text{mod } 11)$, we can begin by cancelling the factor 3 to reach the equivalent congruence $x^2 + 4x - 1 \equiv 0 \,(\text{mod } 11)$. That is,

$$(x+2)^2 - 5 \equiv 0 \,(\text{mod } 11), \text{ or } (x+2)^2 \equiv 5 \equiv 16 \,(\text{mod } 11).$$

Hence $x + 2 \equiv \pm 4 \,(\text{mod } 11)$, which leads to $x \equiv 2 \,(\text{mod } 11)$ and $x \equiv 5 \,(\text{mod } 11)$.

You could, of course, begin by multiplying through by 12, but you should be on the lookout for shortcuts.

(b) The quadratic congruence $2x^2 - 5x + c \equiv 0 \,(\text{mod } 13)$ has discriminant $25 - 8c$, and so the congruence has solutions when this is either a quadratic residue of 13 or congruent to 0 modulo 13. Now $25 - 8c \equiv 5c - 1 \,(\text{mod } 13)$ and, referring to the list of quadratic residues of 13 established in Worked Exercise 1.6, for solutions we require

$$5c - 1 \equiv 0,\ 1,\ 3,\ 4,\ 9,\ 10,\ 12 \,(\text{mod } 13).$$

These linear congruences have, respectively, solutions

$$c \equiv 8,\ 3,\ 6,\ 1,\ 2,\ 10,\ 0 \,(\text{mod } 13).$$

Hence the original quadratic congruence has solutions for these values of c.

Solving for the case $c = 1$,

$$\begin{aligned} 2x^2 - 5x + 1 \equiv 0 \,(\text{mod } 13) &\iff 16x^2 - 40x + 8 \equiv 0 \,(\text{mod } 13) \\ &\iff (4x-5)^2 - 17 \equiv 0 \,(\text{mod } 13) \\ &\iff (4x-5)^2 \equiv 4 \,(\text{mod } 13) \\ &\iff 4x - 5 \equiv \pm 2 \,(\text{mod } 13). \end{aligned}$$

These linear congruences, and hence the quadratic congruence, have solutions $x \equiv 4 \,(\text{mod } 13)$ and $x \equiv 5 \,(\text{mod } 13)$.

(c) (i) $3^{(23-1)/2} = 3^{11} = 27 \times 27 \times 27 \times 9$
$\equiv 4^3 \times 9 \equiv (-5) \times 9 = -45 \equiv 1 \,(\text{mod } 23)$.
So 3 is a quadratic residue of 23.

(ii) $3^{(29-1)/2} = 3^{14} = 27^4 \times 9 \equiv (-2)^4 \times 9$
$= 4 \times 36 \equiv 28 \equiv -1 \,(\text{mod } 29)$.
So 3 is a quadratic non-residue of 29.

(d) Suppose that $n > 2$, $\text{hcf}(a, n) = 1$ and $x^2 \equiv a \,(\text{mod } n)$. Then $\text{hcf}(x^2, n) = 1$, from which it follows that $\text{hcf}(x, n) = 1$, since any common factor of x and n would necessarily divide x^2. Now

$$a^{\phi(n)/2} \equiv (x^2)^{\phi(n)/2} \equiv x^{\phi(n)} \equiv 1 \,(\text{mod } n), \text{ by Euler's Theorem.}$$

A counter-example to the converse is provided by the case $a = 3$ and $n = 8$:

$$a^{\phi(n)/2} = 3^{\phi(8)/2} = 3^{4/2} = 9 \equiv 1 \,(\text{mod } 8).$$

However, the only squares modulo 8 are 0, 1 and 4, and so there is no solution for $x^2 \equiv 3 \,(\text{mod } 8)$.

Solution to Exercise 2.1

$$\begin{aligned}
(30/11) &= (8/11), \quad \text{by property (a)},\\
&= (2/11)(4/11), \quad \text{by property (c)},\\
&= (2/11), \quad \text{by property (b)},\\
&= (-9/11), \quad \text{by property (a)},\\
&= (-1/11)(9/11), \quad \text{by property (c)},\\
&= (-1/11), \quad \text{by property (b)},\\
&= -1, \quad \text{by property (e) or, alternatively, by property (d).}
\end{aligned}$$

Solution to Exercise 2.2

(a) (i) $$\begin{aligned}
(3/11) &= (3/11)(2^2/11), \quad \text{since } (2^2/11) = 1 \text{ by property (b)},\\
&= (3 \times 2^2/11), \quad \text{by property (c)},\\
&= (12/11)\\
&= (1/11), \quad \text{by property (a)},\\
&= 1, \quad \text{by property (b)}.
\end{aligned}$$

(ii) $$\begin{aligned}
(2/17) &= (36/17), \quad \text{by property (a)},\\
&= 1, \quad \text{by property (b), since } 36 = 6^2.
\end{aligned}$$

(iii) $$\begin{aligned}
(40/11) &= (10/11)(4/11), \quad \text{by property (c)},\\
&= (10/11), \quad \text{by property (b)},\\
&= (-1/11), \quad \text{by property (a)},\\
&= -1, \quad \text{by property (e)}.
\end{aligned}$$

(b) For the first part, the essential task is to show that $(2/31) = 1$. There are various ways of achieving this, such as to appeal to Euler's Criterion and to show that $2^{15} \equiv 1 \,(\text{mod } 31)$. Simpler, if you spot it, is to capitalise on the observation that $2 \equiv 64 \,(\text{mod } 31)$ and so $(2/31) = (64/31) = 1$, since $64 = 8^2$. Then

Alternatively use property (d) of Theorem 2.2.

$$(2a/31) = (2/31)(a/31) = (a/31),$$

as required.

To make use of this result, take any quadratic residue of 31 and repeatedly double, finding further quadratic residues. For example, starting from 1:

1, 2, 4, 8 and 16 are quadratic residues.

Starting from $9 = 3^2$, we have further quadratic residues

$$9,\ 18,\ 36 \equiv 5,\ 10,\ 20.$$

Starting from $25 = 5^2$ completes the list:

$$25,\ 50 \equiv 19,\ 38 \equiv 7,\ 14,\ 28.$$

As 31 is known to have fifteen quadratic residues, we have found them all:

$$1, 2, 4, 5, 7, 8, 9, 10, 14, 16, 18, 19, 20, 25 \text{ and } 28.$$

Solution to Exercise 2.3

(a) (i) The discriminant is $8^2 - 4 \times 1 \times 4 = 48$; so our task is to determine the Legendre symbol $(48/13)$.

$$\begin{aligned}(48/13) &= (16/13)(3/13), \quad \text{by property (c)},\\ &= (3/13), \quad \text{by property (b)},\\ &= (16/13), \quad \text{by property (a)},\\ &= 1, \quad \text{by property (b)}.\end{aligned}$$

So the congruence has solutions.

(ii) The discriminant is $7^2 - 4 \times 1 \times 9 = 13$, so our task is to determine the Legendre symbol $(13/23)$.

$$\begin{aligned}(13/23) &= (36/23), \quad \text{by property (a)},\\ &= 1, \quad \text{by property (b)}.\end{aligned}$$

So the congruence has solutions.

The properties are those of Theorem 2.2.

The observation that $(48/13) = (9/13)$, by property (a), suggests an equally good approach.

(b) (i) $$\begin{aligned}(18/19) &= (-1/19), \quad \text{by property (a)},\\ &= -1, \quad \text{by property (e)}.\end{aligned}$$

(ii) $$\begin{aligned}(45/19) &= (5/19)(9/19), \quad \text{by property (c)},\\ &= (5/19), \quad \text{by property (b)},\\ &= (5/19)(4/19), \quad \text{since } (4/19) = 1 \text{ by property (b)},\\ &= (20/19), \quad \text{by property (c)},\\ &= (1/19), \quad \text{by property (a)},\\ &= 1.\end{aligned}$$

(iii) $$\begin{aligned}(2/19) &= (2/19)(9/19), \quad \text{since } (9/19) = 1 \text{ by property (b)},\\ &= (18/19), \quad \text{by property (c)},\\ &= -1, \quad \text{from part (i) above}.\end{aligned}$$

(c) Since $29 \equiv 1 \pmod 4$, $(-1/29) = 1$ by property (e) of Theorem 2.2. It now follows that for any quadratic residue a,

$$(-a/29) = (-1/29)(a/29) = 1 \times 1 = 1,$$

and so $-a$ is also a quadratic residue. In particular, as the 'squares' 1, 4, 9 and 16 are quadratic residues, so too are $-1 \equiv 28$, $-4 \equiv 25$, $-9 \equiv 20$ and $-16 \equiv 13$.

The prime 29 has fourteen quadratic residues and so we seek a further six. Using the fact that 20 is a quadratic residue, we have

$$1 = (20/29) = (4/29)(5/29) = (5/29),$$

There are other ways, such as using larger squares, of finding the remaining quadratic residues.

and so 5, and hence $-5 \equiv 24$, are quadratic residues. Similarly

$$1 = (28/29) = (4/29)(7/29) = (7/29),$$

so that 7 and 22 are quadratic residues. Finally

$$1 = (5/29)(7/29) = (35/29) = (6/29),$$

giving 6 and 23 as the missing pair.

Collecting everything together gives the full list of quadratic residues:

1, 4, 5, 6, 7, 9, 13, 16, 20, 22, 23, 24, 25 and 28.

(d) By property (d) of Theorem 2.2,

$$(a/p) \equiv a^{(p-1)/2} \pmod{p}$$

and therefore

$$(a/p)a^{(p-1)/2} \equiv (a/p)(a/p) \equiv 1 \pmod{p}.$$

Now Wilson's Theorem gives $(p-1)! \equiv -1 \pmod{p}$, and combining the two congruences we have

$$(p-1)! \equiv -(a/p)a^{(p-1)/2} \pmod{p}.$$

Solution to Exercise 3.1

(a) To evaluate (7/13) we first form the set S consisting of the first $\frac{13-1}{2} = 6$ multiples of 7, namely

$$S = \{7, 14, 21, 28, 35, 42\},$$

It is quite possible to go direct to the set S', bypassing S, by starting with 7 and writing down six integers, getting from one to the next by adding 7 modulo 13.

and convert to S' by replacing each integer with its least positive residue modulo 13, to get

$$S' = \{\mathbf{7}, 1, \mathbf{8}, 2, \mathbf{9}, 3\}.$$

The three emboldened members exceed $\frac{13}{2}$ and so

$$(7/13) = (-1)^3 = -1.$$

(b) Repeating for (11/23), we get

$$S = \{11, 22, 33, 44, 55, 66, 77, 88, 99, 110, 121\},$$
$$S' = \{11, \mathbf{22}, 10, \mathbf{21}, 9, \mathbf{20}, 8, \mathbf{19}, 7, \mathbf{18}, 6\}.$$

The five emboldened members of S' exceed $\frac{23}{2}$ and so

$$(11/23) = (-1)^5 = -1.$$

(c) Finally for (5/29), the first fourteen positive multiples of 5 give, modulo 29, the set

$$S' = \{5, 10, \mathbf{15}, \mathbf{20}, \mathbf{25}, 1, 6, 11, \mathbf{16}, \mathbf{21}, \mathbf{26}, 2, 7, 12\},$$

which has the six emboldened members exceeding $\frac{29}{2}$, and so

$$(5/29) = (-1)^6 = 1.$$

Solution to Exercise 3.2

(a) (i) $(2/29) = -1$, since $29 \equiv 5 \pmod 8$.

(ii) $(2/31) = 1$, since $31 \equiv 7 \pmod 8$.

(iii) $(2/43) = -1$, since $43 \equiv 3 \pmod 8$.

(b) By Gauss's Lemma, $(-2/p) = (-1)^n$, where n is the number of integers in the set $S = \{-2, -4, -6, \ldots, -2k, \ldots, -(p-1)\}$ that have least positive residue exceeding $\frac{p}{2}$. The least positive residue modulo p of each of the integers in S is obtained by adding p, giving

$$S' = \{1, 3, \ldots, p-2k, \ldots, p-4, p-2\}.$$

Now $p - 2k > \frac{p}{2}$ if, and only if, $k < \frac{p}{4}$, and therefore $n = \text{int}\left(\frac{p}{4}\right)$. Looking at the four possible congruence classes of p modulo 8:

$$\begin{aligned}
p = 8k+1 &\implies n = \text{int}\left(2k + \tfrac{1}{4}\right) = 2k,\\
p = 8k+3 &\implies n = \text{int}\left(2k + \tfrac{3}{4}\right) = 2k,\\
p = 8k+5 &\implies n = \text{int}\left(2k + \tfrac{5}{4}\right) = 2k+1,\\
p = 8k+7 &\implies n = \text{int}\left(2k + \tfrac{7}{4}\right) = 2k+1.
\end{aligned}$$

Now when n is even, $(-2/p) = 1$, and this occurs when $p \equiv 1$ or $3 \pmod 8$. When n is odd, $(-2/p) = -1$, and this occurs when $p \equiv 5$ or $7 \pmod 8$.

(c) To make use of $(-2/p) = (-1/p)(2/p)$ we look at the four cases for p modulo 8, and multiply the value of $(-1/p)$, obtained from property (e) of Theorem 2.2, by the value of $(2/p)$, obtained from Proposition 3.2.

$$\begin{aligned}
p \equiv 1 \pmod 8 &\implies (-1/p) = 1 \quad \text{and} \quad (2/p) = 1; \quad \text{so } (-2/p) = 1\\
p \equiv 3 \pmod 8 &\implies (-1/p) = -1 \quad \text{and} \quad (2/p) = -1; \quad \text{so } (-2/p) = 1\\
p \equiv 5 \pmod 8 &\implies (-1/p) = 1 \quad \text{and} \quad (2/p) = -1; \quad \text{so } (-2/p) = -1\\
p \equiv 7 \pmod 8 &\implies (-1/p) = -1 \quad \text{and} \quad (2/p) = 1; \quad \text{so } (-2/p) = -1
\end{aligned}$$

This agrees with the result obtained in part (b).

Solution to Exercise 3.3

Suppose that $p_1, p_2, \ldots, p_n$ are the only primes of form $8k+3$, and let $N = (p_1 p_2 \cdots p_n)^2 + 2$. Since, for each $1 \le i \le n$, $p_i^2 \equiv 1 \pmod 8$, we note that $N \equiv 3 \pmod 8$. Now if p is any prime factor of the odd integer N we have $(p_1 p_2 \cdots p_n)^2 \equiv -2 \pmod p$ and so, as the left-hand side is a square, $(-2/p) = 1$. The result of Exercise 3.2(b) confirms that p is congruent modulo 8 either to 1 or to 3.

In the solution to Worked Exercise 3.9 of Chapter 9 we saw that the square of any odd integer, prime or composite, is congruent modulo 8 to 1.

At this point we can invoke the argument seen in the proof of Proposition 3.5. If all the prime factors of N are congruent modulo 8 to 1 then so too is N itself. But N is not congruent modulo 8 to 1 and so it must have at least one prime factor, say q, that is congruent modulo 8 to 3. Hence q must be one of the listed primes p_i. But then q divides N and q divides $(p_1 p_2 \cdots p_n)^2$, which leads to the contradiction that q divides 2.

Hence there are infinitely many primes of the form $8k+3$.

Solution to Exercise 3.4

(a) (i) The set S comprising the first $\frac{13-1}{2}$ multiples of -3 is

$$S = \{-3, -6, -9, -12, -15, -18\};$$

and replacing each integer by its least positive residue modulo 13,

$$S' = \{\mathbf{10}, \mathbf{7}, 4, 1, \mathbf{11}, \mathbf{8}\}.$$

The four emboldened members exceed $\frac{13}{2}$ and so $(-3/13) = (-1)^4 = 1$.

(ii) Repeating for $(5/17)$:

We have gone directly to S' by starting with 5 and moving from one number to the next by adding 5 modulo 17 until we have eight numbers.

$$S' = \{5, \mathbf{10}, \mathbf{15}, 3, 8, \mathbf{13}, 1, 6\}.$$

Three members of S' exceed $\frac{17}{2}$ and so $(5/17) = (-1)^3 = -1$.

(b) Using Proposition 3.2, which tells us that 2 is a quadratic residue of p if, and only if, $p \equiv \pm 1 \pmod 8$, we aim to determine the value of each of the given primes modulo 8.

(i) For $n > 1$, 8 divides 4^n and so $4^n + 1 \equiv 1 \pmod 8$.

(ii) For $p \geq 3$, 8 divides 2^p and so $2^p - 1 \equiv -1 \pmod 8$. However, for $p = 2$, $M_2 = 3$ and hence 2 is not a quadratic residue of M_2.

(iii) For $n > 1$, 8 divides 2^{2^n} and so $2^{2^n} + 1 \equiv 1 \pmod 8$.

In all cases 2 is a quadratic residue of the given prime.

From $(-2/p) = (2/p)(-1/p)$ it follows that -2 will be a quadratic residue provided -1 is a quadratic residue, and this we know, from property (e) of Theorem 2.2, will occur exactly when $p \equiv 1 \pmod 4$. Hence we have the following results.

(i) $4^n + 1 \equiv 1 \pmod 4$, and so -2 is a quadratic residue.

Notice that -2 is a quadratic residue of $M_2 = 3$.

(ii) For $p > 2$, $2^p - 1 \equiv -1 \pmod 4$, and so -2 is a quadratic non-residue.

(iii) $2^{2^n} + 1 \equiv 1 \pmod 4$, and so -2 is a quadratic residue.

(c) Suppose that $p_1, p_2, \ldots, p_n$ are the only primes of form $8k + 5$. Consider the integer $N = 4(p_1 p_2 \cdots p_n)^2 + 1$. We observe that as each $p_i^2 \equiv 25 \equiv 1 \pmod 8$, N itself is of the form $8k + 5$. Being odd, N has some odd prime factor p, and the congruence

$$4(p_1 p_2 \cdots p_n)^2 \equiv -1 \pmod p$$

leads to $(-1/p) = 1$, as the left-hand side is a square. Property (e) of Theorem 2.2 now gives $p \equiv 1 \pmod 4$; that is, $p \equiv 1 \pmod 8$ or $p \equiv 5 \pmod 8$.

The rest of the argument should be familiar. If all the prime factors of N are of the form $8k + 1$ then so too would be N itself. This is not the case and hence N must have at least one prime factor of the form $8k + 5$; that is, N is divisible by one of the primes p_i. But then p_i divides $N - 4(p_1 p_2 \cdots p_n)^2 = 1$, which is the required contradiction.

Hence there are infinitely many primes of the form $8k + 5$.

Solution to Exercise 4.1

(a) (i) There are many possible sequences of steps, of which this is one. We have given an explanation of each step rather than a reference to the result being used.

$$\begin{aligned}
(-219/383) &= (-1/383)(3/383)(73/383), \quad \text{since } -219 = (-1)\times 3\times 73,\\
&= (-1)(3/383)(73/383), \quad \text{since } 383 \equiv 3 \,(\text{mod } 4),\\
&= (-1)(-1)(383/3)(73/383), \quad \text{by LQR since } 383 \equiv 3 \,(\text{mod } 4),\\
&= (2/3)(73/383), \quad \text{since } 383 \equiv 2 \,(\text{mod } 3),\\
&= (-1)(73/383), \quad \text{since } 3 \equiv 3 \,(\text{mod } 8),\\
&= (-1)(383/73), \quad \text{by LQR since } 73 \equiv 1 \,(\text{mod } 4),\\
&= (-1)(18/73), \quad \text{since } 383 \equiv 18 \,(\text{mod } 73),\\
&= (-1)(2/73)(9/73), \quad \text{since } 18 = 2\times 9,\\
&= (-1)(1)(9/73), \quad \text{since } 73 \equiv 1 \,(\text{mod } 8),\\
&= (-1)(1)(1), \quad \text{since } 9 = 3^2,\\
&= -1.
\end{aligned}$$

We have consistently adopted the 'invert and divide' approach in this solution, but there are other methods. For example, the observation that $-219 \equiv 164 \,(\text{mod } 383)$ and $164 = 4\times 41$ offers

$$(-219/383) = (4/383)(41/383) = (41/383) = \cdots$$

as an alternative start.

(ii) Whereas we spelled out each step in part (i), here we reduce the number of steps by processing several symbols simultaneously.

$$\begin{aligned}
(461/773) &= (773/461), \quad \text{by LQR},\\
&= (312/461), \quad \text{since } 773 \equiv 312 \,(\text{mod } 461),\\
&= (4/461)(2/461)(3/461)(13/461), \quad \text{since } 312 = 2^3\times 3\times 13,\\
&= (1)(-1)(461/3)(461/13), \quad \text{since } 461 \equiv 5 \,(\text{mod } 8) \text{ and by LQR twice},\\
&= (-1)(2/3)(6/13), \quad \text{since } 461 \equiv 2 \,(\text{mod } 3) \text{ and } 461 \equiv 6 \,(\text{mod } 13),\\
&= (-1)(-1)(2/13)(3/13), \quad \text{since } (2/3) = -1 \text{ and } 6 = 2\times 3,\\
&= (-1)(13/3), \quad \text{since } 13 \equiv 5 \,(\text{mod } 8) \text{ and by LQR},\\
&= (-1)(1/3), \quad \text{since } 13 \equiv 1 \,(\text{mod } 3),\\
&= -1.
\end{aligned}$$

(b)
$$\begin{aligned}
(164/197) &= (4/197)(41/197), \quad \text{since } 164 = 4\times 41,\\
&= (41/197), \quad \text{since 4 is a square},\\
&= (197/41), \quad \text{by LQR},\\
&= (74/41), \quad \text{since } 197 \equiv 74 \,(\text{mod } 41),\\
&= (2/41)(37/41), \quad \text{since } 74 = 2\times 37,\\
&= (37/41), \quad \text{since } 41 \equiv 1 \,(\text{mod } 8),\\
&= (41/37), \quad \text{by LQR},\\
&= (4/37), \quad \text{since } 41 \equiv 4 \,(\text{mod } 37),\\
&= 1, \quad \text{since 4 is a square}.
\end{aligned}$$

(c) There are three major mistakes in the argument.

1. The step that replaces $(-19/79)$ with $(-1)(19/79)$ turns out to be correct but is handling the minus sign incorrectly. It should read $(-19/79) = (-1/79)(19/79)$. However, it happens that $(-1/79) = -1$, since $79 \equiv 3 \pmod 4$, so the misuse has not resulted in a wrong value being carried forward.
2. The first application of LQR is wrong. As $19 \equiv 79 \equiv 3 \pmod 4$, LQR gives $(19/79) = (-1)(79/19)$.
3. The second application of LQR is also wrong. As 15 is not a prime, LQR cannot be applied to Legendre symbol $(15/19)$. The correct continuation here would be $(15/19) = (3/19)(5/19)$. This mistake is compounded in the next line too, where the symbol $(19/15)$ is manipulated. As 15 is not prime this is not a legitimate Legendre symbol.

Solution to Exercise 4.2

(a) The quadratic congruence $3x^2 + 6x + 5 \equiv 0 \pmod{89}$ is solvable provided that $6^2 - 4 \times 3 \times 5 = -24$ is a quadratic residue of the prime 89.

$$\begin{aligned}(-24/89) &= (65/89), \quad \text{since } -24 \equiv 65 \pmod{89},\\ &= (5/89)(13/89), \quad \text{using } 65 = 5 \times 13,\\ &= (89/5)(89/13), \quad \text{by LQR, since } 89 \equiv 1 \pmod 4,\\ &= (4/5)(11/13), \quad \text{since } 89 \equiv 4 \pmod 5 \text{ and } 89 \equiv 11 \pmod{13},\\ &= (1)(13/11), \quad \text{since } 4 = 2^2 \text{ and by LQR, since } 13 \equiv 1 \pmod 4,\\ &= (2/11), \quad \text{since } 13 \equiv 2 \pmod{11},\\ &= -1, \quad \text{since } 11 \equiv 3 \pmod 8.\end{aligned}$$

Hence the congruence has no solutions.

(b) The quadratic congruence $2x^2 + 5x - 9 \equiv 0 \pmod{11 \times 101}$ has solutions if, and only if, each of the congruences $2x^2 + 5x - 9 \equiv 0 \pmod{11}$ and $2x^2 + 5x - 9 \equiv 0 \pmod{101}$ are solvable. This will be the case if $5^2 - 4 \times 2 \times (-9) = 97$ is a quadratic residue of both 11 and 101.

See Subsection 4.3 of Book A, Chapter 4.

Note that 97 and 101 are primes.

$$\begin{aligned}(97/11) &= (9/11) = (3^2/11) = 1\\ (97/101) &= (101/97), \quad \text{by LQR, since } 101 \equiv 1 \pmod 4,\\ &= (4/97) = (2^2/97) = 1\end{aligned}$$

Hence the congruence does have solutions.

Solution to Exercise 4.3

(a) We make use of the property $(6/p) = (2/p)(3/p)$ and the established facts

$$\begin{aligned}(2/p) &= 1 \quad \text{if, and only if, } p \equiv \pm 1 \pmod 8;\\ (3/p) &= 1 \quad \text{if, and only if, } p \equiv \pm 1 \pmod{12}.\end{aligned}$$

To cater for modulo 8 and modulo 12 we will consider the possible values of p modulo 24, which (since $p > 3$) are 1, 5, 7, 11, 13, 17, 19 and 23. $\text{lcm}(8, 12) = 24$

$$\begin{aligned}
p \equiv 1 \,(\text{mod } 24) &\implies (2/p) = 1 \quad \text{and} \quad (3/p) = 1, \quad \text{so } (6/p) = 1\\
p \equiv 5 \,(\text{mod } 24) &\implies (2/p) = -1 \quad \text{and} \quad (3/p) = -1, \quad \text{so } (6/p) = 1\\
p \equiv 7 \,(\text{mod } 24) &\implies (2/p) = 1 \quad \text{and} \quad (3/p) = -1, \quad \text{so } (6/p) = -1\\
p \equiv 11 \,(\text{mod } 24) &\implies (2/p) = -1 \quad \text{and} \quad (3/p) = 1, \quad \text{so } (6/p) = -1\\
p \equiv 13 \,(\text{mod } 24) &\implies (2/p) = -1 \quad \text{and} \quad (3/p) = 1, \quad \text{so } (6/p) = -1\\
p \equiv 17 \,(\text{mod } 24) &\implies (2/p) = 1 \quad \text{and} \quad (3/p) = -1, \quad \text{so } (6/p) = -1\\
p \equiv 19 \,(\text{mod } 24) &\implies (2/p) = -1 \quad \text{and} \quad (3/p) = -1, \quad \text{so } (6/p) = 1\\
p \equiv 23 \,(\text{mod } 24) &\implies (2/p) = 1 \quad \text{and} \quad (3/p) = 1, \quad \text{so } (6/p) = 1
\end{aligned}$$

So $(6/p) = 1$ when $p \equiv 1$, 5, 19 or 23 (mod 24).

(b) We will use the property $(-3/p) = (-1/p)(3/p)$ and the established facts

$(-1/p) = 1$ if, and only if, $p \equiv 1 \,(\text{mod } 4)$;
$(3/p) = 1$ if, and only if, $p \equiv \pm 1 \,(\text{mod } 12)$.

We need to consider possible values of p modulo 12, which, since $p > 3$, are 1, 5, 7 and 11.

$$\begin{aligned}
p \equiv 1 \,(\text{mod } 12) &\implies (-1/p) = 1 \quad \text{and} \quad (3/p) = 1, \quad \text{so } (-3/p) = 1\\
p \equiv 5 \,(\text{mod } 12) &\implies (-1/p) = 1 \quad \text{and} \quad (3/p) = -1, \quad \text{so } (-3/p) = -1\\
p \equiv 7 \,(\text{mod } 12) &\implies (-1/p) = -1 \quad \text{and} \quad (3/p) = -1, \quad \text{so } (-3/p) = 1\\
p \equiv 11 \,(\text{mod } 12) &\implies (-1/p) = -1 \quad \text{and} \quad (3/p) = 1, \quad \text{so } (-3/p) = -1
\end{aligned}$$

So $(-3/p) = 1$ when $p \equiv 1$ or 7 (mod 12) and $(-3/p) = -1$ when $p \equiv 5$ or 11 (mod 12). That is,

$$(-3/p) = \begin{cases} 1, & \text{if } p \equiv 1 \,(\text{mod } 6),\\ -1, & \text{if } p \equiv 5 \,(\text{mod } 6), \end{cases}$$

as claimed.

Now suppose that $p_1, p_2, \ldots, p_n$ are the only primes of the form $6k + 1$. Consider $N = (2p_1p_2 \cdots p_n)^2 + 3$. As N is odd it has an odd prime factor, say p. Then $(2p_1p_2 \cdots p_n)^2 \equiv -3 \,(\text{mod } p)$, from which it follows that $(-3/p) = 1$, and the result above gives $p \equiv 1 \,(\text{mod } 6)$. It now follows that p is one of the listed primes and so p is a factor of $(2p_1p_2 \cdots p_n)^2$, and therefore of $N - (2p_1p_2 \cdots p_n)^2$. That is, p divides 3, and so $p = 3$. This contradicts the fact that p is of the form $6k + 1$. It follows that the assumption that there are only finitely many primes of this form must be false.

Therefore there are infinitely many primes of the form $6k + 1$.

Solution to Exercise 4.4

The properties are those of Theorem 2.2.

(a) (i) $(35/71) = (5/71)(7/71)$, by property (c),
$= (71/5)(-1)(71/7)$, by LQR, noting $7 \equiv 71 \equiv 3 \pmod 4$,
$= (1/5)(-1)(1/7)$, by property (a),
$= -1$.

(ii) $(139/149) = (149/139)$, by LQR, noting $149 \equiv 1 \pmod 4$,
$= (10/139)$, by property (a),
$= (2/139)(5/139)$, by property (c),
$= (-1)(139/5)$, by Proposition 3.2 and LQR respectively,
$= (-1)(4/5)$, by property (a),
$= -1$, by property (b).

(iii) $(1234/4567) = (2/4567)(617/4567)$, by property (c),
$= (1)(4567/617)$, by Proposition 3.2 and LQR, noting that 617 is prime and $617 \equiv 1 \pmod 4$,
$= (248/617)$, by property (a) since $7 \times 617 = 4319$,
$= (4/617)(2/617)(31/617)$, by property (c),
$= (1)(1)(617/31)$, by property (b), Proposition 3.2 and LQR,
$= (28/31)$, by property (a),
$= (4/31)(7/31)$, by property (c),
$= (1)(-1)(31/7)$, by property (b) and LQR,
$= (-1)(3/7)$, by property (a),
$= (-1)(-1)(7/3)$, by LQR,
$= (1/3)$, by property (a),
$= 1$.

(b) (i) The discriminant is $1^2 - 4 \times 3 \times (-7) = 85 \equiv 0 \pmod{17}$, and so the congruence has a unique solution.

(ii) The discriminant is $1^2 - 4 \times 2 \times 1 = -7$, and so the existence or otherwise of solutions is determined by the value of $(-7/29)$.

$(-7/29) = (-1/29)(7/29)$, by property (c),
$= (1)(29/7)$, by property (e) and LQR,
$= (1/7) = 1$.

The congruence has solutions.

(iii) The discriminant is $3^2 - 4 \times 2 \times (-7) = 65$, and so we investigate $(65/101)$.

$$\begin{aligned}(65/101) &= (5/101)(13/101), \quad \text{by property (c)},\\ &= (101/5)(101/13), \quad \text{by LQR},\\ &= (1/5)(10/13), \quad \text{by property (a)},\\ &= (1)(10/13)\\ &= (40/13), \quad \text{by property (c), using } (4/13) = 1,\\ &= (1/13) = 1.\end{aligned}$$

The congruence has solutions.

(c) (i) If $p = 4k + 3$ then

$$q = 10p + 3 = 40k + 33 = 4(10k + 8) + 1 = 4k' + 1.$$

From this we see that p and q cannot both be congruent modulo 4 to 3. Hence LQR gives

$$(p/q) = (q/p) = (10p + 3/p) = (3/p).$$

Note that $10p + 3 \equiv 3 \pmod{p}$.

(ii) We consider separately the possible values of p modulo 4.

If $p = 4k + 1$ then

$$(-1/p) = 1 \text{ and } (p/q) = (q/p) = (10p + 1/p) = (1/p) = 1.$$

If $p = 4k + 3$ then

$$(-1/p) = -1 \text{ and } q = 10p + 1 = 4(10k + 7) + 3,$$

and so

$$(p/q) = (-1)(q/p) = (-1)(10p + 1/p) = (-1)(1/p) = -1.$$

In both cases $(p/q) = (-1/p)$.

(d) The fact that $4^n \equiv 4 \pmod{12}$, for all $n \geq 1$, is readily proved by mathematical induction. It is certainly true for $n = 1$ and the assumption that $4^k \equiv 4 \pmod{12}$ leads to

$$4^{k+1} = 4^k \times 4 \equiv 4 \times 4 \equiv 4 \pmod{12}.$$

Using this we obtain the following.

(a) $4^n + 1 \equiv 5 \pmod{12}$, for all $n \geq 1$.

(b) $2^p - 1 = 2 \times 4^{(p-1)/2} - 1 \equiv 2 \times 4 - 1 \equiv 7 \pmod{12}$, for all $p \geq 3$.

(c) $2^{2^n} + 1 \equiv 4^{2^{n-1}} + 1 \equiv 5 \pmod{12}$, for all $n \geq 1$.

By Proposition 4.7, the integer 3 is a quadratic non-residue of each of these primes.

(e) For odd primes $p \neq 7$ we have

$$(7/p) = \begin{cases} (p/7), & \text{if } p \equiv 1 \pmod 4, \\ (-1)(p/7), & \text{if } p \equiv 3 \pmod 4. \end{cases}$$

$1^2 = 1,\ 2^2 = 4,\ 3^2 \equiv 2 \pmod 7$

The quadratic residues of 7 are 1, 2 and 4, and so

$$(p/7) = \begin{cases} 1, & \text{if } p \equiv 1, 2 \text{ or } 4 \pmod 7, \\ -1, & \text{if } p \equiv 3, 5 \text{ or } 6 \pmod 7. \end{cases}$$

Putting the two parts together, $(7/p) = 1$ if, and only if,

$$\begin{array}{rlll} \text{either} & p \equiv 1 \pmod 4 & \text{and} & p \equiv 1, 2 \text{ or } 4 \pmod 7, \\ \text{or} & p \equiv 3 \pmod 4 & \text{and} & p \equiv 3, 5 \text{ or } 6 \pmod 7. \end{array}$$

That is,

$(7/p) = 1$ if, and only if, $p \equiv 1, 3, 9, 19, 25 \text{ or } 27 \pmod{28}$.

(f) Consider first the case $p \equiv q \equiv 3 \pmod 4$. Then

$$(-p/q) = (-1/q)(p/q) = (-1)(p/q) = (-1)(-1)(q/p) = (q/p)$$

while

$$(-q/p) = (-1/p)(q/p) = (-1)(q/p),$$

and so $(-p/q) \neq (-q/p)$.

In all other cases

$$(-p/q) = (-1/q)(p/q) = (-1/q)(q/p),$$

while

$$(-q/p) = (-1/p)(q/p),$$

and the two will be equal when $(-1/q) = (-1/p)$. This happens when $p \equiv q \pmod 4$. As the case $p \equiv q \equiv 3 \pmod 4$ has been eliminated, we conclude that

$(-p/q) = (-q/p)$ if, and only if, $p \equiv q \equiv 1 \pmod 4$.

Chapter 11

Rings and polynomials

Introduction

In Book A and Chapters 9 and 10, we studied properties of the set $\mathbb{Z}$ of integers. In this chapter and the next, we are going to take a step into the abstract to discover what 'factorisation' means for the elements of a set, equipped with a suitable definition for multiplication.

In this chapter, we will first explore the abstract definitions of a *ring* and a *field*, before studying *polynomials* where the coefficients of the polynomials are elements from a given field. It turns out that these 'polynomial rings' have many properties in common with the integers; for example we can define a division algorithm, and this enables us to develop the analogue of the 'highest common factor' for two polynomials. In the final section of this chapter, we will look more closely at polynomials with integer coefficients, and prove several results relating to the factorisation of a more general collection of polynomials, namely those with *rational* coefficients.

Chapter guide

This chapter has four sections, with each section relying, to an extent, on the content of the previous sections. The first section contains relatively few proofs, but is quite long and defines algebraic concepts such as rings and fields, which are required by the rest of this chapter as well as Chapter 12 and much of Book E. The second section is shorter than the other three, and defines concepts needed for the final two sections, although these themes also reoccur in later chapters. The last two sections are not built upon as much by later chapters as the first two, although the concepts they contain do reoccur, particularly in Chapter 12.

1 Rings

In Book B we learnt about groups, which are defined as a set of elements equipped with some operation $\circ$, and satisfying certain axioms given in Definition 1.1 of Book B, Chapter 5. When we think of the set of integers, however, there are two basic operations that we can use: addition, $+$, and multiplication, $\cdot$. We learnt in Chapter 5 that $(\mathbb{Z}, +)$ is an abelian group, but we can quickly verify that $(\mathbb{Z}, \cdot)$ is not: no integers apart from 1 and -1 have multiplicative inverses and so axiom G3 of Definition 1.1 does not hold.

What about subtraction and division? In a sense, we can think of these as the inverses to addition and multiplication.

We may ask which of the group axioms *do* hold for $(\mathbb{Z}, \cdot)$. If $m, n \in \mathbb{Z}$, then $m \cdot n \in \mathbb{Z}$ and so we have closure (axiom G1). Next, since $1 \cdot m = m \cdot 1 = m$, the element $1 \in \mathbb{Z}$ acts as the identity to confirm axiom G2. Finally, associativity of multiplication is a basic property of the integers and so G4 holds. The integers form a model for our definition of a ring.

The old German word *Ring* can mean 'association'; hence the terms 'ring' and 'group' have similar origins.

Definition 1.1 *Ring axioms*

Let R be a set and let $+$ and $\cdot$ be binary operations defined on R. Then $(R, +, \cdot)$ is a **ring** if the following axioms hold.

Axioms for addition:

R1 Closure For all $a, b \in R$,

$$a + b \in R.$$

R2 Associativity For all $a, b, c \in R$,

$$a + (b + c) = (a + b) + c.$$

R3 Additive identity There exists an additive identity $0 \in R$ such that, for all $a \in R$,

$$a + 0 = a = 0 + a.$$

R4 Additive inverses For each $a \in R$, there exists an additive inverse $-a \in R$ such that

$$a + (-a) = 0 = (-a) + a.$$

R5 Commutativity For all $a, b \in R$,

$$a + b = b + a.$$

Axioms for multiplication:

R6 Closure For all $a, b \in R$,

$$a \cdot b \in R.$$

R7 Associativity For all $a, b, c \in R$,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

R8 Multiplicative identity There exists a multiplicative identity $1 \in R$ such that, for all $a \in R$,

$$a \cdot 1 = a = 1 \cdot a.$$

Axioms combining addition and multiplication:

R9 Distributive laws For all $a, b, c \in R$, multiplication is left distributive over addition in R:

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c),$$

and multiplication is right distributive over addition in R:

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c).$$

Furthermore, $(R, +, \cdot)$ is a **commutative ring** if the following extra axiom holds.

R10 Commutativity For all $a, b \in R$,

$$a \cdot b = b \cdot a.$$

At first sight, this looks like a lot of individual axioms to satisfy; so let us look at them collectively. First off, note that the axioms for addition (R1 to R5) are exactly the axioms required for $(R, +)$ to be an abelian group. As a result, we can use the 'Elementary consequences' (see, for example, Proposition 1.2 of Book B, Chapter 5) to deduce the following.

Lemma 1.2

Let $(R, +, \cdot)$ be a ring. Then the additive identity $0 \in R$ is unique. Furthermore, for every $a \in R$ the additive inverse $(-a)$ is unique.

Next, R6 to R8 tell us some rules about multiplication, but note that it is not the case that $(R, \cdot)$ is a group, because we are missing the axiom that guarantees the existence of multiplicative inverses. Finally, R9 shows us how to combine these two operations, in exactly the way we expect when multiplying and adding integers.

Early definitions of a ring did not include the multiplicative identity axiom, and this approach persists in some texts. In this case, the ring we have defined above is called a **ring with 1**, to distinguish it from a 'ring without 1'. However, all our rings will have a multiplicative identity, and so we can use the term 'ring' unambiguously.

Most of the rings in this module will also be commutative (that is, axiom R10 holds), but as we are about to see an example of a non-commutative ring, we will not for the moment assume this axiom.

Example 1.3 *Some familiar rings*

(a) $(\mathbb{Z}, +, \cdot)$, the set of integers $\mathbb{Z}$, with normal addition and multiplication, is a commutative ring.

This was our original 'template' to create the abstract definition of a ring. So it is appropriate that it is the first example we encounter.

To see this, first note that $(\mathbb{Z}, +)$ forms an infinite abelian group, with additive identity 0. This, therefore, immediately covers axioms R1 to R5. Similarly, R6 to R8 follow since multiplication over the integers is closed, associative, and the multiplicative identity is 1. Of course, multiplication over the integers is commutative, and so we can also verify axiom R10. This leaves only axiom R9, which follows as distributivity is a basic property of the integers.

(b) $(\mathbb{Q}, +, \cdot)$, the set of rational numbers, with normal addition and multiplication, forms a commutative ring. Similarly, we can readily check that the set of real numbers, $\mathbb{R}$, and the set of complex numbers, $\mathbb{C}$, are also both commutative rings under the same operations.

(c) $(M_{2\times 2}, +, \cdot)$, the set of 2×2 matrices over $\mathbb{R}$, with matrix addition and multiplication, is a ring but not a commutative ring.

Checking some of the axioms for this example takes a little more thought. For addition, closure follows since the sum of two 2×2 matrices is another 2×2 matrix. It can easily be checked that the addition of matrices is associative and commutative. The additive

identity is the zero matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and consequently the additive inverse of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\begin{pmatrix} -a & -b \\ -c & -d \end{pmatrix}$.

For multiplication, when we multiply two 2×2 matrices together we obtain another 2×2 matrix, and so closure (R6) is satisfied. It is straightforward (but not immediately obvious) to check that multiplication is associative, and we can check that the identity matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ acts as the multiplicative identity for axiom R8.

Finally, axiom R9 can be readily verified, but it is well known that multiplication of matrices is not commutative and so R10 does not hold. For example:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \text{ but } \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}.$$

Exercise 1.1

The ring $\mathbb{Z} \times \mathbb{Z}$ consists of all ordered pairs of integers (a, b), with addition and multiplication defined by $(a, b) + (c, d) = (a + c, b + d)$ and $(a, b) \cdot (c, d) = (ac, bd)$. Write down:

(a) $(2, 1) + (4, 5)$ and $(2, 1) \cdot (4, 5)$

(b) the additive identity element

(c) the additive inverse of the element $(2, 1)$

(d) the multiplicative identity element.

Exercise 1.2

Show that the following are commutative rings.

(a) The set of integers $\mathbb{Z}_n = \{0, 1, 2, \ldots, n - 1\}$, equipped with addition and multiplication modulo n.

(b) The set

$$\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}\},$$

equipped with the usual addition and multiplication of real numbers.

You may be wondering how we justify statements such as 'multiplication of integers is associative'. The properties of integers that you have known (not necessarily by name) and used since you first started doing arithmetic can be justified only by working with a formal definition of the integers. A well-known definition of the integers starts from a set of axioms known as 'Peano axioms', after Giuseppe Peano (1858–1932). These are then extended to give 'Peano arithmetic'. These axioms include the properties

such as associativity, commutativity and distributivity that we have been discussing, but in this module you are not expected to be familiar with these axioms; the approach we have used so far to justify such statements is sufficient.

In light of the examples above, you may now be wondering what objects *do not* form rings; so let us look at a couple before we proceed.

Example 1.4 *Two examples of non-rings*

(a) The set of non-negative integers, $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$, with the usual addition and multiplication.

We find that R1 and R2 are true, but then we cannot find additive inverses to satisfy R3: for example, we need -1 to be the additive inverse of 1, but it is not in the set $\mathbb{Z}^+$.

(b) The even integers, $2\mathbb{Z} = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$, with the usual addition and multiplication.

In this example, we find that this set does actually form an abelian group under addition. The problem comes when we try to find a multiplicative identity for axiom R8: there isn't one! If there was a multiplicative identity, $2k \in 2\mathbb{Z}$, say, then $(2k) \cdot (2n) = 2n$ for all $n \in \mathbb{Z}$. This, however, implies that $2k = 1$, and so $k \notin \mathbb{Z}$, which is a contradiction.

Having seen some examples and non-examples, we are ready to prove some simple properties that we should expect to find. Throughout these, you should be thinking how this is exactly what we expect in the examples of rings that we have just seen. Are the properties also true for each of our two non-examples, and if they aren't, why not?

Proposition 1.5 *Basic properties of rings*

Let $(R, +, \cdot)$ be a ring. Then:

(a) $0 \cdot a = a \cdot 0 = 0$ for all $a \in R$

(b) $-(-a) = a$ for all $a \in R$

(c) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$ for all $a, b \in R$

(d) $(-a) \cdot (-b) = a \cdot b$ for all $a, b \in R$.

Proving statements like these is actually quite tricky: we need to be careful that at each stage we use only the axioms that define a ring, or results that we have derived from these axioms.

Proof We will prove parts (a)–(c), and leave part (d) as an exercise for you to do.

(a) Let $a \in R$, and set $b = 0 \cdot a$. We want to show that b is the additive identity. First, we know that $0 = 0 + 0$, and so $0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$, using axiom R9. Since we set $b = 0 \cdot a$, we have $b = b + b$. Therefore, using associativity:

$$0 = b + (-b) = (b + b) + (-b) = b + (b + (-b)) = b + 0 = b.$$

A similar argument can be used to prove that $a \cdot 0 = 0$.

(b) First, for any $a \in R$ we have $(-a) + a = 0$. Now, the element $(-a)$ of R has a unique additive inverse (by Lemma 1.2), namely $-(-a)$, from which it follows that $a = -(-a)$.

(c) We have $a \cdot b + (-a) \cdot b = (a + (-a)) \cdot b = 0 \cdot b = 0$, the last step following by part (a) of the proposition. Therefore, $(-a) \cdot b$ must be the additive inverse of $a \cdot b$, and so we have $(-a) \cdot b = -(a \cdot b)$. Similarly, we can show that $-(a \cdot b) = a \cdot (-b)$. ■

Having established these basic properties, we will relax our notation a little. We have so far used the minus sign '$-$' to denote the additive inverse of an element: $a + (-a) = 0$. When working with the integers, we are familiar with using the minus sign as a binary operation, that is, we write expressions such as '$a - b$'. We will extend this notation here. Formally, in any ring R with $a, b \in R$, we define $a - b$ to mean the result of adding the additive inverse of b to a. Put simply, $a - b = a + (-b)$.

When the context is clear, we will use two further notational simplifications, similar to those made at the start of Book B. First, we will write the juxtaposition ab to mean the product $a \cdot b$. Second, we will often refer to $(R, +, \cdot)$ by the set symbol R when it is clear which operations are being used.

Exercise 1.3

Prove part (d) of Proposition 1.5.

Exercise 1.4

What can you say about a ring in which the additive and multiplicative identities are the same, that is, a ring where $0 = 1$?

1.1 Subrings

In Exercise 1.2, we showed that the sets $\mathbb{Z}_n$ and $\mathbb{Z}[\sqrt{2}]$ were rings. In the solution, we appealed to properties of addition and multiplication of a larger set to prove some of the axioms. For example, $\mathbb{Z}[\sqrt{2}]$ is a subset of the ring $\mathbb{R}$, and hence, since we are using the same definition of addition and multiplication for both, some of the axioms hold immediately. In this subsection, we want to formalise this approach.

Definition 1.6 *Subring*

Let R be a ring. We say that S is a **subring** of R if:

(a) the set S is a subset of R

(b) S is a ring when equipped with the same binary operations as R

(c) the ring S has the same multiplicative identity as R.

Why do we insist that S has the same multiplicative identity? As we will see shortly in Worked Exercise 1.9(b), it is possible for a subset of a ring to be a ring itself, but with a different multiplicative identity. We want to be able to distinguish between this case and the definition given above.

Why, then, did we not also insist that S should have the same additive identity as R? We can deduce this from the definition given above.

Lemma 1.7

Let R be a ring, and S a subring of R. Then the additive identity of S is the same as the additive identity of R.

Proof Suppose 0_S is the additive identity in S, and 0_R that for R. Then for any $s \in S$, we have $s + 0_S = s$ as an equation in S, and so also as an equation in R. Working in R, we now have:

$$\begin{aligned} 0_R &= (-s) + s \\ &= (-s) + (s + 0_S) \\ &= ((-s) + s) + 0_S = 0_R + 0_S \\ &= 0_S. \end{aligned}$$ ■

Some obvious examples are the integers $\mathbb{Z}$ as a subring of the rationals $\mathbb{Q}$, and the rationals as a subring of the reals $\mathbb{R}$, and the reals as a subring of the complex numbers $\mathbb{C}$. We saw in Example 1.3 that all of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ are rings, and for each the additive identity is 0 and the multiplicative identity is 1.

Now, as we wanted, we can state and prove a lemma justifying why we do not need to check all the axioms R1–R9 when verifying that a subset of a ring is a subring.

Lemma 1.8 *Subring criterion*

Let R be a ring and let S be a subset of R. Then S is a subring of R if:

SR1 for all $s, t \in S$, $s - t \in S$

SR2 for all $s, t \in S$, $st \in S$

SR3 the multiplicative identity of R is in S.

Axiom SR1 can be thought of as 'closure under subtraction'.

Proof Since S is a subset of R, we just need to check that S is a ring with the same binary operations and multiplicative identity as R.

Let $s \in S$. (Note that S is non-empty by SR3; so such an s exists.) By SR1, $s - s = 0 \in S$ and so $0 - s = (-s) \in S$. In addition, if $t \in S$ then $(-t) \in S$, so that $s + t = s - (-t) \in S$. This shows that $(S, +)$ is an abelian subgroup of $(R, +)$.

Multiplication is closed in S by SR2, and axiom SR3 ensures that the multiplicative identity of R is in S, and in fact it must be a multiplicative identity for S. The associative and distributive laws hold since they hold in R. Thus we have shown that S is a ring, and since the multiplicative identity coincides, it follows immediately that S is a subring of R. ■

Worked Exercise 1.9

(a) Show that the set of **Gaussian integers**, $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$ where $i = \sqrt{-1}$, equipped with the usual addition and multiplication of complex numbers, is a ring.

(b) Why is $S = \{0, 5\}$ not a subring of $\mathbb{Z}_{10}$?

Solution

(a) Since $\mathbb{Z}[i]$ is a subset of the complex numbers $\mathbb{C}$, it will be enough to prove that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. We check the conditions of Lemma 1.8. Consider two elements $a + ib$ and $c + id$ of $\mathbb{Z}[i]$. Then

SR1: $(a + ib) - (c + id) = (a - c) + i(b - d) \in \mathbb{Z}[i]$.

SR2: $(a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc) \in \mathbb{Z}[i]$.

SR3: The multiplicative identity of $\mathbb{C}$ and $\mathbb{Z}[i]$ is $1 = 1 + 0i$.

It follows from the subring criterion that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$, and therefore that $\mathbb{Z}[i]$ is a ring.

(b) Condition SR3 does not hold for S: the multiplicative identity of $\mathbb{Z}_{10}$ is 1, which is not in S.

Note, however, that S is a ring with the same operations as $\mathbb{Z}_{10}$. The multiplicative identity of S is 5.

The requirement that the multiplicative identity of a subring is always the same as that of the original ring has some interesting consequences. For example, the only subring of $\mathbb{Z}$ is $\mathbb{Z}$ itself, and the only subring of $\mathbb{Z}_n$ is $\mathbb{Z}_n$.

Exercise 1.5

By considering a suitable larger ring and using Lemma 1.8, prove that the following are rings:

(a) $\mathbb{Z}[\omega] = \{a + b\omega + c\omega^2 : a, b, c \in \mathbb{Z}\}$ where $\omega = e^{2i\pi/3}$, so that $\omega^3 = 1$

(b) $\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$.

1.2 Zero divisors and units

The axioms that define a ring leave some opportunities for results that we might not expect. By considering the integers $\mathbb{Z}$, we already know that elements do not need to have multiplicative inverses, but in some rings even stranger things can happen.

Suppose that in $\mathbb{Z}$ we wish to solve the quadratic equation

$$x^2 - 5x + 6 = 0.$$

Factorising the left-hand side, we obtain

$$(x - 2)(x - 3) = 0.$$

It follows that either $x - 2 = 0$ or $x - 3 = 0$, and hence that $x = 2$ or $x = 3$. This method works since if in $\mathbb{Z}$ the product of two terms is 0, then one of the two terms must itself be 0.

More generally, if we know that the product of two integers a and b is 0, then we can deduce that at least one of a or b must be 0. Put another way, if we multiply two non-zero integers then their product must also be non-zero.

However, in the ring $\mathbb{Z}_6$ we have $2 \cdot 3 \equiv 0 \,(\mathrm{mod}\, 6)$. In other words, there are two non-zero elements whose product is zero. We want to be able to distinguish between elements where this does happen and those where it does not, and so we make the following definition.

Definition 1.10 *Zero divisor*

Let R be a ring. A non-zero element $a \in R$ is a **zero divisor** in R if there is a non-zero element $b \in R$ for which $ab = 0$ or $ba = 0$.

In a sense, 0 is also a zero divisor since $0 \cdot 1 = 0$ is true in every ring, but we have deliberately excluded this so that in future we do not have to keep using phrases such as 'no non-zero zero divisors'. Commutative rings with no zero divisors have a special name. Since they have several of the properties of the ring of integers, we call them **integral domains**.

We will learn more about integral domains in Chapter 12.

Example 1.11 *Rings with no zero divisors*

(a) The ring $\mathbb{Z}$ has no zero divisors – this is the fact that we used to solve the quadratic equation above. Similarly, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ also have no zero divisors.

(b) $\mathbb{Z}_7$ has no zero divisors. If $ab \equiv 0 \pmod 7$, then 7 divides ab, and hence (since 7 is prime) divides a or b (or both) in $\mathbb{Z}_7$. Thus $a = 0$ or $b = 0$.

Example 1.12 *Rings with zero divisors*

(a) $\mathbb{Z}_6$, as we saw above, does have zero divisors. Since $2 \cdot 3 \equiv 0 \pmod 6$ and $3 \cdot 4 \equiv 0 \pmod 6$, we see that all of 2, 3 and 4 are zero divisors.

However, 1 and 5 are not zero divisors since there are no numbers a and b (other than 0) in $\mathbb{Z}_6$ for which $1 \cdot a \equiv 0 \pmod 6$ or $5 \cdot b \equiv 0 \pmod 6$.

(b) The ring $M_{2\times 2}$ of 2×2 matrices with entries from $\mathbb{R}$ has zero divisors. For example, the product

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

shows that both $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ are zero divisors.

Exercise 1.6

Let R be the ring $\mathbb{Z} \times \mathbb{Z}$ (as defined in Exercise 1.1). For $a, b, c, d \in \mathbb{Z}$, addition is defined by $(a, b) + (c, d) = (a + c, b + d)$, and multiplication is given by $(a, b) \cdot (c, d) = (ac, bd)$.

(a) Give an example of a zero divisor in R.

(b) Describe all the zero divisors in R.

Now that we are aware of the possibility that rings have zero divisors, we turn again to the issue of multiplicative inverses: which elements of a ring can be multiplied together to give 1? For example, in $\mathbb{Z}$, we have $(-1) \cdot (-1) = 1$, and so -1 is its own multiplicative inverse, but there is no multiplicative inverse of 2 in $\mathbb{Z}$.

So far, we have allowed our rings to be non-commutative, that is, axiom R10 has not needed to be satisfied. Our main example of a non-commutative ring has been $M_{2\times 2}$, the ring of 2×2 matrices with entries from $\mathbb{R}$. As we move our focus to consider elements that have multiplicative inverses, the existence of non-commutative rings leads to some awkward possibilities. For example, could we have a non-commutative ring R and $a, b \in R$, with $a \cdot b = 1$, but $b \cdot a \neq 1$?

The answer to this question is 'yes' for some quite peculiar rings, but we will not explore this here.

Convention: commutative rings

To avoid the complications described above, all rings we will consider from now on will be commutative. Additionally, we will drop the word 'commutative' and simply use the term 'ring'.

With this settled, we can define a term for elements that have multiplicative inverses.

Definition 1.13 *Unit*

An element a of a (commutative) ring R is a **unit** if there is an element $a^{-1} \in R$ for which $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

The term 'unit' is overused in mathematics generally. It should be clear from the context which meaning is appropriate.

For example, in the ring $\mathbb{Z}$ the only units are 1 and -1; in the ring $\mathbb{Z}_{10}$ the units are 1, 3, 7 and 9, since

$$1 \times 1 \equiv 1 \,(\text{mod } 10),\ 3 \times 7 \equiv 1 \,(\text{mod } 10) \text{ and } 9 \times 9 \equiv 1 \,(\text{mod } 10).$$

Exercise 1.7

List all the units in the following rings.

(a) $\mathbb{Z}_{12}$

(b) $\mathbb{Z} \times \mathbb{Z}$

(c) $\mathbb{Z}[i]$

Although in the above exercise it was relatively straightforward to list the units in specific rings, this is not always the case. We will see examples in Chapter 12 where finding units is rather more difficult.

Exercise 1.8

(a) Write down all the zero divisors in the rings $\mathbb{Z}_n$, for $n = 8$, 9, 10, 11 and 12.

(b) Give a description of all the zero divisors in $\mathbb{Z}_n$, for any given natural number n.

(c) Which rings $\mathbb{Z}_n$ have no zero divisors?

(d) Show that in a ring R, if a is a unit then it cannot be a zero divisor.

1.3 Fields

In the definition of a ring, the axiom we were missing for the non-zero elements to form a group under multiplication was that every element has an inverse: in other words, every non-zero element is a unit. A ring with this property is called a *field*.

Definition 1.14 *Field*

A (commutative) ring R is a **field** if the additive and multiplicative identities are distinct, and every non-zero element is a unit.

In other words, R has distinct elements 0 and 1, axioms R1–R10 hold, and so does:

R11 Multiplicative inverses For every non-zero $a \in R$, there exists $a^{-1} \in R$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Recall Exercise 1.4, which tells us what happens when $0 = 1$ in a ring.

In the same way that we have been using the symbol R to refer to a general ring, we will use F for a field. You may find that in other texts the letter K is used: this is because the German mathematician Richard Dedekind (1831–1916) used the term *Körper* for field, which translates as 'body'.

Example 1.15 *Some well-known fields*

The following are easily seen to be fields.

(a) $\mathbb{Q}$: Since $\frac{p}{q} \cdot \frac{q}{p} = 1$ for non-zero $p, q \in \mathbb{Z}$, the inverse of $\frac{p}{q}$ is $\frac{q}{p}$.

(b) $\mathbb{R}$: The inverse of any non-zero $x \in \mathbb{R}$ is $x^{-1} = \frac{1}{x}$.

(c) $\mathbb{C}$: The inverse of $z = a + ib \neq 0$ is

$$z^{-1} = \frac{1}{a+ib} = \left(\frac{a}{a^2+b^2}\right) + i\left(\frac{-b}{a^2+b^2}\right).$$

Example 1.16 *Some rings that are not fields*

(a) $\mathbb{Z}$ is not a field: as we observed in the previous subsection, the only units are -1 and 1, and so no other integer has a multiplicative inverse.

(b) $\mathbb{Z}_6$ is not a field: for example, the element 2 is not a unit.

Since $\mathbb{Z}_6$ has sufficiently few elements, we can verify this by testing all multiples of 2 modulo 6:

$$2 \cdot 0 \equiv 0, \quad 2 \cdot 1 \equiv 2, \quad 2 \cdot 2 \equiv 4,$$
$$2 \cdot 3 \equiv 0, \quad 2 \cdot 4 \equiv 2, \quad 2 \cdot 5 \equiv 4.$$

Exercise 1.9 *Properties of multiplicative inverses*

Let F be a field, and $a, b \in F$ non-zero elements. Prove the following.

(a) $(a^{-1})^{-1} = a$

(b) $(ab)^{-1} = b^{-1}a^{-1}$

The second example of a non-field, $\mathbb{Z}_6$, raises an interesting point: we used the element 2, which is a zero divisor, since $2 \cdot 3 \equiv 0 \pmod 6$. In Exercise 1.8(d), we showed that any element that is a unit cannot also be a zero divisor. As a consequence of this, we have the following lemma.

Lemma 1.17

Let F be a field. Then F has no zero divisors.

Proof Suppose $a \neq 0$ is a zero divisor in F, and so there exists a non-zero $b \in F$ such that $a \cdot b = 0$. By axiom R11, there exists $b^{-1} \in F$ such that $b \cdot b^{-1} = 1$. Now

$$a = a \cdot 1 = a \cdot (b \cdot b^{-1}) = (a \cdot b) \cdot b^{-1} = 0 \cdot b^{-1} = 0,$$

which is a contradiction. ■

In Exercise 1.8, we stated that $\mathbb{Z}_n$ has no zero divisors if, and only if, n is a prime number. In fact, each element of $\mathbb{Z}_n$ is either a zero divisor or a unit.

Lemma 1.18

Let $a \in \mathbb{Z}_n$ be non-zero, where $n \in \mathbb{N}$. Then a is either a zero divisor or a unit.

Proof If $\text{hcf}(a, n) = 1$, then by Proposition 4.7 of Book A, Chapter 1 there exist $s, t \in \mathbb{Z}$ such that $sa + tn = 1$. Therefore, $sa \equiv 1 \pmod n$, and hence a is a unit in $\mathbb{Z}_n$.

On the other hand, if $\text{hcf}(a, n) = d > 1$, then $a = dh$ and $n = dk$ for some h, k, with $1 < k < n$ and $\text{hcf}(h, k) = 1$. Then

$$ak = (dh)k = (hd)k = h(dk) = hn \equiv 0 \pmod n,$$

which proves that a is a zero divisor. ■

For any non-zero element a, one of two things can happen: either $\text{hcf}(a, n) = 1$, or $\text{hcf}(a, n) > 1$. We consider each of these possibilities in turn.

This lemma immediately gives us an important consequence.

Corollary 1.19

The ring $\mathbb{Z}_n$ is a field if, and only if, n is a prime number.

The fields $\mathbb{Z}_p$, where p is a prime number, contain only finitely many elements, and are therefore called **finite fields**. We will not explore finite fields further in this chapter, as we will study them in some detail in Book E, Chapter 19. However, we will make two quick remarks.

First, the fields $\mathbb{Z}_p$, for p prime, are not the only finite fields: in fact, there exists a finite field with q elements if, and only if, q is a power of a prime number.

Second, if we repeatedly add the multiplicative identity to itself, $1 + 1 + \cdots + 1$, we will at some point end up with the additive identity. For example, in $\mathbb{Z}_3$, $1 + 1 + 1 = 3 \equiv 0 \pmod 3$. The number of times we add 1 to itself to get 0 is known as the **characteristic**. So, the characteristic of $\mathbb{Z}_3$ is 3.

We finish this section by defining subfields, and developing an analogue to the subring criterion (Lemma 1.8) that allows us to avoid checking all the axioms.

Definition 1.20 *Subfield*

A subset S of a field F that is a field with the same binary operations and multiplicative identity as F is known as a **subfield** of F.

Lemma 1.21 *Subfield criterion*

Let F be a field, and let S be any subset of F. Then S is a subfield of F if:

(a) S is a subring of F

(b) every non-zero element of S has a multiplicative inverse in S.

Consequently, to determine whether S is a subfield of F or not, we need to check only axioms SR1–SR3 from Lemma 1.8 and axiom R11 from Definition 1.14.

Proof Since S is a subset of F, we just need to check that S is a field with the same binary operations and identities as F. First, condition (a) tells us that S has the same multiplicative identity as F, and that all the axioms R1–R10 hold. This leaves only axiom R11 about the existence of multiplicative inverses, but this is guaranteed by condition (b). ■

Exercise 1.10

By considering suitable larger fields, show that the following are fields.

(a) $\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$

(b) $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$

2 Polynomials over fields

As we saw in Subsection 1.2, rings can have unexpected properties. To avoid some of these issues, we are going to spend the remainder of this chapter studying polynomial rings where the coefficients of the polynomials come from a field, as these are generally much better behaved than generic rings. There will be one exception to this rule: towards the end of Section 4 we will encounter the polynomial ring $\mathbb{Z}[x]$, even though $\mathbb{Z}$ is not a field.

2.1 Polynomial rings

Definition 2.1 *Polynomial ring over a field*

Let F be a field. Then a **polynomial over** F with the variable x is a polynomial of the form $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, where $a_0, a_1, a_2, \ldots, a_n \in F$ and $n \geq 0$, where x^0 is defined to be 1.

The **polynomial ring over** F is

$$F[x] = \{a_0 + a_1x + \cdots + a_nx^n : a_0, a_1, \ldots, a_n \in F, n \geq 0\}.$$

We have seen the square bracket notation '$F[x]$' a few times already in this chapter, and it is worth verifying that our use of it has been consistent. For example, in Exercise 1.10(b) we encountered $\mathbb{Q}[\sqrt{2}]$, which was defined to be the set $\{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$. The above definition would define it as $\{a_0 + a_1\sqrt{2} + \cdots + a_n(\sqrt{2})^n : a_i \in \mathbb{Q}\}$, but of course $(\sqrt{2})^2 = 2$ is already an element of $\mathbb{Q}$, and so the higher terms of $\sqrt{2}$ can simply be included in the coefficients a_0 and a_1.

To be clear that $F[x]$ is indeed a ring, we need to define 'addition' and 'multiplication' within $F[x]$, and then be sure that all the axioms hold. How we add and multiply polynomials should come as no surprise: first, we need to remember that the field F came with its own addition and multiplication, and use these to build the natural definitions. If in doubt, think about how it is done in $\mathbb{R}[x]$, that is, when the coefficients are all real numbers.

Addition is defined by

$$\left(\sum_{i=0}^{n} a_ix^i\right) + \left(\sum_{i=0}^{m} b_ix^i\right) = \sum_{i=0}^{\max(m,n)} (a_i + b_i)x^i.$$

Note that the sum on the right-hand side of this equation goes up to $\max(m, n)$: so, if $n < m$ then we add 'dummy' coefficients $a_{n+1} = a_{n+2} = \cdots = a_m = 0$.

What is the zero of $F[x]$? If '0' is the symbol for the zero in the field F, then let 0_x be the zero in $F[x]$. It is defined as the polynomial where every coefficient is equal to 0; so in fact we have $0_x = 0$. However, we will

continue to use the symbol 0_x to highlight when we are working in the polynomial ring $F[x]$.

We can now work out additive inverses:

$$\left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{n} (-a_i) x^i\right) = \sum_{i=0}^{n} \left(a_i + (-a_i)\right) x^i = 0_x.$$

The closure, associativity and commutativity of addition in $F[x]$ follow quite easily: remember that we can assume they hold in F.

Next, we define multiplication:

$$\left(\sum_{i=0}^{n} a_i x^i\right) \cdot \left(\sum_{j=0}^{m} b_j x^j\right) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i \cdot b_j\right) x^k.$$

At first sight, this definition looks rather offputting. So, let us quickly consider an example to reassure ourselves that this is exactly what we expect to see. Take polynomials $f(x) = 1 + 5x - 4x^2$ and $g(x) = 2 + 3x$ over $\mathbb{R}$. Then

$$\begin{aligned} f(x)g(x) &= (1 + 5x - 4x^2) \cdot (2 + 3x) \\ &= \big(1 \cdot 2\big) + \big(1 \cdot 3 + 5 \cdot 2\big)x + \big(5 \cdot 3 + (-4) \cdot 2\big)x^2 + \big((-4) \cdot 3\big)x^3 \\ &= 2 + 13x + 7x^2 - 12x^3. \end{aligned}$$

What is the multiplicative identity of $F[x]$? As we did with the zero, we will use the special symbol 1_x to distinguish it from the '1' in F, even though $1_x = 1$.

The remaining axioms R6–R9 are readily verified, as a consequence of the axioms for the field F. In addition, since multiplication in F is commutative, it follows that $F[x]$ is a commutative ring, that is, axiom R10 holds.

Example 2.2 *Some common polynomial rings*

The polynomial ring $\mathbb{Q}[x]$ will appear in Section 4.

(a) The set of polynomials with rational coefficients, $\mathbb{Q}[x]$, with $0_x = 0$ and $1_x = 1$, is a ring since $\mathbb{Q}$ is a field.

(b) $\mathbb{C}[x]$ represents the polynomials with complex coefficients, and $\mathbb{R}[x]$ the polynomials with real coefficients. Note that $\mathbb{Q}[x] \subset \mathbb{R}[x] \subset \mathbb{C}[x]$.

(c) The polynomial ring $\mathbb{Z}_2[x]$ consists of all polynomials whose coefficients are all either 0 or 1. More generally, we can consider the polynomial ring $\mathbb{Z}_p[x]$ where p is a prime number.

Worked Exercise 2.3

Let $f(x) = 1 + 2x$ and $g(x) = 1 + 3x + 2x^3$ be polynomials in $\mathbb{Z}_5[x]$, the ring of polynomials over the field $\mathbb{Z}_5$, equipped with addition and multiplication modulo 5.

(a) Calculate and simplify (i) $f(x) + g(x)$, and (ii) $3f(x) - g(x)$.

(b) Compute $f(x)g(x)$.

Solution

(a) (i) $f(x) + g(x) = (1 + 1) + (2 + 3)x + 2x^3 \equiv 2 + 2x^3 \pmod 5$

(ii)
$$\begin{aligned} 3f(x) - g(x) &= 3(1 + 2x) - (1 + 3x + 2x^3) \\ &= (3 + (3 \times 2)x) - 1 - 3x - 2x^3 \\ &\equiv (3 + x) + 4 + 2x + 3x^3 \\ &= (3 + 4) + (1 + 2)x + 3x^3 \\ &\equiv 2 + 3x + 3x^3 \pmod 5 \end{aligned}$$

(b)
$$\begin{aligned} f(x)g(x) &= (1 + 2x)(1 + 3x + 2x^3) \\ &= 1 + 5x + 6x^2 + 2x^3 + 4x^4 \\ &\equiv 1 + x^2 + 2x^3 + 4x^4 \pmod 5 \end{aligned}$$

Note that it is sometimes easier when working in $\mathbb{Z}_p[x]$ to calculate the polynomials over $\mathbb{Z}$, and then reduce the coefficients modulo p at the end.

Exercise 2.1

Let $f(x) = 2x^3 - 3$ and $g(x) = 9x^4 + 7x^2 - x + 1$. Compute (i) $f(x) + g(x)$ and (ii) $f(x) \cdot g(x)$ in the following polynomial rings:

(a) $\mathbb{Q}[x]$

(b) $\mathbb{Z}_7[x]$

(c) $\mathbb{Z}_2[x]$.

2.2 The degree of a polynomial

Definition 2.4 *Degree of a polynomial*

Let F be a field, and let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial in $F[x]$. The **degree** of $f(x)$, $\deg(f)$, is the largest $k \geq 0$ for which $a_k \neq 0$.

This non-zero coefficient a_k is known as the **leading coefficient**.

At this point, it is also worth making a remark about notation: strictly speaking, we should have written $\deg(f(x))$ rather than $\deg(f)$, but this is rather clumsy. For the same reason, in future we will occasionally drop the '(x)' from other equations involving polynomials, where the meaning is still sufficiently clear.

Example 2.5 *The degree of a polynomial*

In Exercise 2.1, the polynomial $f(x) = 2x^3 - 3$ has degree 3 in $\mathbb{Q}[x]$ and $\mathbb{Z}_7[x]$, but $f(x) \equiv 1 \pmod 2$ and so $f(x)$ has degree 0 in $\mathbb{Z}_2[x]$.

On the other hand, the polynomial $g(x) = 9x^4 + 7x^2 - x + 1$ has degree 4 in all of $\mathbb{Q}[x]$, $\mathbb{Z}_7[x]$ and $\mathbb{Z}_2[x]$.

Convention: degree of the zero polynomial

By convention, the degree of the polynomial 0_x will be $-\infty$, which should be interpreted as meaning we can choose the degree of 0_x to be as largely negative as we need.

This choice is made so that the following lemma is consistent when one or both of the polynomials f and g are equal to 0_x.

Lemma 2.6

Let F be a field, and let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Then $\deg(fg) = \deg(f) + \deg(g)$.

This seemingly innocuous lemma actually turns out to be quite important, as we will see later in this chapter and the next. Before we prove it, let us pause to think what would happen if we replaced the field F with some arbitrary ring R. We consider an example in the following exercise.

Exercise 2.2

Consider the polynomials $f(x) = 2x^2 + x + 3$ and $g(x) = 3x + 1$ in the polynomial ring $\mathbb{Z}_6[x]$. Find:

(a) $\deg(f)$

(b) $\deg(g)$

(c) $\deg(fg)$.

What went wrong? We expected the degree of $f(x)g(x)$ to be 3, but the coefficient of x^3 disappeared because $2 \cdot 3 \equiv 0 \pmod 6$. This, of course, is because 2 and 3 are zero divisors in $\mathbb{Z}_6$, and we know from Lemma 1.17 that fields cannot have zero divisors.

Proof of Lemma 2.6 First, if either $f(x)$ or $g(x)$ is equal to 0_x, then $fg = 0_x$ and $\deg(fg) = -\infty$ as we should expect. So we now suppose that $f(x) \neq 0_x$ and $g(x) \neq 0_x$.

Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$ be polynomials over F, where $\deg(f) = n$ and $\deg(g) = m$. In particular, this means that $a_n \neq 0$ and $b_m \neq 0$. Now, from the definition of multiplication in $F[x]$,

$$f(x)g(x) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i \cdot b_j \right) x^k.$$

The highest term in this sum is $a_nb_mx^{m+n}$, and so we certainly have $\deg(fg) \leq \deg(f) + \deg(g)$. Moreover, since $a_n, b_m \in F$ are both not equal to zero, it follows by Lemma 1.17 that their product a_nb_m is also not 0. Thus $\deg(fg) = m + n = \deg(f) + \deg(g)$. ■

Meanwhile, what happens to the degree if we add two polynomials? Clearly, from the definition, $\deg(f + g) \leq \max(\deg(f), \deg(g))$, but can we replace the '$\leq$' with '$=$'? Again, let us first consider a couple of examples.

Exercise 2.3

Calculate the degree of $f(x) + g(x)$ over the stated ring:

(a) $f(x) = 2x^3 + 4x^2 - 1$ and $g(x) = 3x^3 + 3x^2 - 2x$ over $\mathbb{Z}_5$

(b) $f(x) = 1 + x$ and $g(x) = 1 - x$ over $\mathbb{Q}$.

In the above exercise, you should have found that it is indeed possible for the degree of the sum of two polynomials to be less than the sum of the degrees. In each case, the top term of $f + g$ disappeared because $\deg(f) = \deg(g)$, and the leading coefficients of f and g were additive inverses of each other.

Lemma 2.7

Let F be a field, and let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Then $\deg(f + g) \leq \max(\deg(f), \deg(g))$.

Proof This follows directly from the definition of addition for polynomials in $F[x]$. ■

2.3 Basic properties of polynomial rings over fields

Now we have defined and developed the concept of the degree of a polynomial, we are in a position to prove some results to reassure us that polynomials over fields behave in the way that we expect them to.

First, the property that the field F has no zero divisors can be shown to imply that there are also no zero divisors in $F[x]$.

Lemma 2.8

Let F be a field, and let $f(x), g(x) \in F[x]$ be polynomials such that $f(x)g(x) = 0_x$. Then either $f(x) = 0_x$, or $g(x) = 0_x$ (or both).

Proof Since by Lemma 2.6 we have $-\infty = \deg(0_x) = \deg(fg)$ $= \deg(f) + \deg(g)$, it follows that at least one of $f(x)$ or $g(x)$ must have degree $-\infty$, that is, $f(x) = 0_x$ or $g(x) = 0_x$. ■

We might now wonder whether in fact $F[x]$ is a field: is every element a unit? This turns out not to be the case.

Lemma 2.9

An element of a polynomial ring $F[x]$ over a field F has a multiplicative inverse (that is, the element is a unit) if, and only if, it has degree 0.

Proof First, a polynomial $f(x)$ of degree 0 is simply a non-zero element of the field F: $f(x) = a_0$ with $a_0 \in F$. Since F is a field, there exists $a_0^{-1} \in F$ such that $a_0 \cdot a_0^{-1} = 1$. Therefore, the inverse of $f(x)$ in $F[x]$ is $f^{-1}(x) = a_0^{-1}$.

Conversely, for a polynomial $f(x) \in F[x]$, suppose that $g(x) \in F[x]$ is the multiplicative inverse. Then $f(x)g(x) = 1_x$, and so by Lemma 2.6 $\deg(f) + \deg(g) = \deg(1_x) = 0$. By definition, the degree of a polynomial is either $-\infty$ or an integer value greater than or equal to 0. So the only possible situation in which $\deg(f) + \deg(g) = 0$ arises is when f has degree 0. ■

Since every unit in a polynomial ring $F[x]$ is actually in F, we will typically refer to $u \in F[x]$ rather than $u(x) \in F[x]$ from now on.

Lemma 2.9 implies that $F[x]$ is not itself a field, since we cannot find inverses for anything with degree ≥ 1. However, despite this lack of inverses in $F[x]$, we can still cancel terms from equations.

Lemma 2.10 *Cancellation of polynomials*

If $f(x)$ is a non-zero polynomial and $f(x)g(x) = f(x)h(x)$ in $F[x]$, then $g(x) = h(x)$.

Proof We can rearrange the equation $fg = fh$ to obtain $fg - fh = 0$, which by distributivity gives us $f(g - h) = 0$. Now $g - h$ is a polynomial over F, so by Lemma 2.8 either $f(x) = 0_x$ (which by assumption is not true), or $g(x) - h(x) = 0_x$. Therefore, we must have $g(x) = h(x)$. ■

We finish this section with a couple of definitions that we will need later in this chapter.

Definition 2.11 *Associate and monic*

Let F be a field, and $f(x), g(x) \in F[x]$.

(a) We say that $f(x)$ is an **associate** of $g(x)$ if $f(x) = ug(x)$ where u is a unit in $F[x]$.

(b) The polynomial $f(x)$ is **monic** if the leading coefficient is equal to 1.

We will generalise the concept of 'associate' to other rings in Chapter 12.

Exercise 2.4

Let F be a field, and $f(x)$ a non-zero polynomial in $F[x]$. Prove the following.

(a) If $g(x) \in F[x]$ is an associate of $f(x)$, then $\deg(g) = \deg(f)$.

(b) There exists a unique monic polynomial that is an associate of $f(x)$.

3 Divisibility of polynomials

This section seeks to develop some results concerning the divisibility of polynomials that mirror results in Book A, Chapter 1 for the integers.

3.1 Polynomial division

You may already have had practice at dividing one polynomial over the real numbers by another and calculating the remainder. Here, we want to extend this 'long division' to polynomials over an arbitrary field F. Before we do this, however, we should justify that what we are doing is valid, and can be done in a unique way: what we need is a division algorithm for polynomials, to mirror Theorem 4.1 in Book A, Chapter 1. In fact, as we proceed through this section you may like periodically to glance back at Section 4 of Chapter 1, and observe the similarities.

Theorem 3.1 *Division Algorithm for polynomials*

Let F be a field and let f and g be polynomials in $F[x]$, with $g \neq 0_x$. Then there exist unique polynomials q and r in $F[x]$ such that

(a) $\deg(r) < \deg(g)$, and

(b) $f = qg + r$.

The polynomials q and r are respectively known as the **quotient** and **remainder** when dividing f by g.

There are two parts to this proof. First, we prove the existence of a suitable q and r using induction, and then we prove that they are unique.

Proof If $f = 0_x$ or $\deg(f) < \deg(g)$, then we can simply take $q = 0_x$ and $r = f(x)$. Thus from now on we will assume $\deg(f) \geq \deg(g)$. Let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ and $g(x) = b_0 + b_1x + \cdots + b_mx^m$, with $a_n \neq 0$ and $b_m \neq 0$ so that $\deg(f) = n$ and $\deg(g) = m$. We use induction on $d = n - m \geq 0$ to prove the existence of a suitable $q(x)$ and $r(x)$.

Remember F is a field, and so we can find b_m^{-1} such that $b_m b_m^{-1} = 1$.

The base case is $d = 0$, when $m = n$. In this case, we can take $q = a_n b_m^{-1}$, and $r = f - qg$. Note that $\deg(r) < \deg(g)$ because the coefficient of x^m in r disappears: $a_m - a_m b_m^{-1} b_m = 0$.

We have expressly constructed f_1 so that $\deg(f_1) < \deg(f)$.

For the inductive step, suppose that the statement is true for $d = 0, 1, \ldots, k-1$. Now consider $d = k$. Define $f_1 = f - a_n b_m^{-1} x^k g$. There are three cases to consider: $f_1 = 0_x$, $\deg(f_1) < \deg(g)$ and $\deg(f_1) \geq \deg(g)$. If $f_1 = 0_x$, then we take $q = a_n b_m^{-1} x^{n-m}$ and $r = 0_x$. If $\deg(f_1) < \deg(g)$, then we take $q = a_n b_m^{-1} x^{n-m}$ and $r = f_1$. Thus we now suppose $\deg(f_1) \geq \deg(g)$. By induction, we can find $q_1(x), r_1(x)$ such that $f_1 = q_1 g + r_1$, with $\deg(r_1) < \deg(g)$. We now substitute back:

$$\begin{aligned} f &= a_n b_m^{-1} x^{n-m} g + f_1 \\ &= a_n b_m^{-1} x^{n-m} g + (q_1 g + r_1) \\ &= \left(a_n b_m^{-1} x^{n-m} + q_1\right) g + r_1 \end{aligned}$$

and so we take $q = a_n b_m^{-1} x^{n-m} + q_1$ and $r = r_1$.

For uniqueness of q and r, suppose that we can find another pair of polynomials q^*, r^* with $f = q^* g + r^*$. Then we have $qg + r = q^* g + r^*$, which can be rearranged to give $(r - r^*) = (q^* - q)g$. Applying Lemma 2.6, we get $\deg(r - r^*) = \deg(q^* - q) + \deg(g)$, but since $\deg(g) > \deg(r - r^*)$ this is a contradiction unless $r = r^*$ and $q^* = q$. ■

Actually finding q and r, when working in a polynomial ring, requires long division of polynomials. We will work through the steps of one example, in case you are not familiar with the process or need reminding.

Worked Exercise 3.2

Find the quotient and remainder on dividing $f(x) = x^4 + x^2 + x - 2$ by $g(x) = x^2 - 1$ in $\mathbb{Q}[x]$.

Solution

We want to find $q(x)$ and $r(x)$ with $f(x) = g(x)q(x) + r(x)$ and the degree of $r(x)$ smaller than that of $g(x)$ (so that $q(x)$ has as high a degree as possible).

First, we write the terms of $f(x)$ and $g(x)$ in descending powers of x, leaving a 'gap' for any powers of x that are missing in $f(x)$.

$$\begin{array}{r|l} x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

Next, take the highest term of $f(x)$, which is x^4, and divide it by the highest term of $g(x)$, which is x^2. The result is the first term, x^2, of the quotient $q(x)$, which we write above the line.

$$\begin{array}{r|l} & x^2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

Now, multiply $g(x)$ by this first term of $q(x)$, and write the terms below the corresponding terms of $f(x)$. Then subtract these terms from $f(x)$.

$$\begin{array}{r|l} & x^2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \\ & x^4 \quad - \quad x^2 \\ \hline & \qquad 2x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

We repeat the process with this new polynomial, $2x^2 + x - 2$, in place of $f(x)$: take the highest term of the new polynomial, which is $2x^2$, and divide it by the highest term of $g(x)$, x^2, to get the next term, 2, of $q(x)$.

$$\begin{array}{r|l} & x^2 \quad + \quad 2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \\ & x^4 \quad - \quad x^2 \\ \hline & \qquad 2x^2 \quad + \quad x \quad - \quad 2 \end{array}$$

Then multiply $g(x)$ by this new term of $q(x)$, and subtract from the polynomial at the bottom.

$$\begin{array}{r|l} & x^2 \quad + \quad 2 \\ x^2 - 1 & x^4 \quad + \quad x^2 \quad + \quad x \quad - \quad 2 \\ & x^4 \quad - \quad x^2 \\ \hline & \qquad 2x^2 \quad + \quad x \quad - \quad 2 \\ & \qquad 2x^2 \qquad \quad - \quad 2 \\ \hline & \qquad\qquad x \end{array}$$

We would now repeat the process, except that the highest term of $g(x)$ is now larger than the highest term of the polynomial we want to divide, and hence what is left is the remainder. This gives $q(x) = x^2 + 2$ and $r(x) = x$.

Exercise 3.1

For each of the following polynomials $f(x), g(x)$ in $\mathbb{Q}[x]$, find the quotient $q(x)$ and the remainder $r(x)$ for the division of $f(x)$ by $g(x)$.

(a) $f(x) = x^3 - 2x^2 + 3x - 1,\ g(x) = x - 1$

(b) $f(x) = 2x^4 - x + 1,\ g(x) = x^2 + 1$

(c) $f(x) = 3x^3 - 2x^2 + 1,\ g(x) = 2x + 1$

With a division algorithm in hand, we now state precisely what we mean when we say that one polynomial divides another: the remainder in the above process must equal the zero polynomial 0_x in $F[x]$.

Definition 3.3 *Factors of a polynomial*

Let F be a field, and f, g polynomials in $F[x]$. We say that g **divides** f (or g is a **factor** of f) if there is some polynomial $h \in F[x]$, such that $f = gh$.

If f is non-zero and neither g nor h are units then g and h are **proper factors** of f.

Notice that if g and h are proper factors of f in $F[x]$, then $0 < \deg(g) < \deg(f)$ and $0 < \deg(h) < \deg(f)$: if $\deg(g) = 0$ then $g \in F$ would be a unit, while if $\deg(g) = \deg(f)$ then Lemma 2.6 of this chapter would imply $\deg(h) = 0$, and so h would be a unit.

Once again, the zero polynomial 0_x behaves slightly differently: every polynomial is a factor of 0_x since $0_x = f(x) \cdot 0_x$. The following exercise will enable you to develop properties of polynomial division that are directly analogous to those we proved for the integers in Section 4 of Book A, Chapter 1.

Exercise 3.2 *Properties of division for polynomials*

Let F be a field, and let f, g and h be non-zero polynomials in $F[x]$. Prove the following statements.

(a) If g divides f then g divides $f + gh$ for any polynomial $h \in F[x]$.

(b) If h divides g and g divides f then h divides f.

(c) If h divides f and h divides g then h divides $af + bg$ for any polynomials $a, b \in F[x]$.

(d) The polynomials f and g are associates if, and only if, g divides f and f divides g.

3.2 Highest common factors

We have now laid the foundations we need to enable us to describe the analogue of the Euclidean Algorithm for polynomials, and hence develop the concept of highest common factor. We proceed in exactly the same way as we did for the integers in Book A, Chapter 1.

Definition 3.4 *Highest common factor of two polynomials*

Let F be a field, and f, g two polynomials in $F[x]$, not both equal to 0_x. Then the **highest common factor** of f and g, written $\text{hcf}(f, g)$, is a monic polynomial of largest degree satisfying the following:

(a) $\text{hcf}(f, g)$ divides both f and g.

(b) Any polynomial $d \in F[x]$ that divides both f and g must also divide $\text{hcf}(f, g)$.

As with the highest common factor of two integers, the hcf of two polynomials is unique, as the next result will show. First, however, we should ask ourselves whether the condition that $\text{hcf}(f, g)$ is monic is strictly necessary. It is straightforward to see that yes, it is necessary to ensure uniqueness; otherwise, any associate of $\text{hcf}(f, g)$ would also satisfy conditions (a) and (b) of Definition 3.4.

Theorem 3.5

Let F be a field, and f, g two polynomials in $F[x]$, not both equal to 0_x. Then $\text{hcf}(f, g)$ is unique, and there exist polynomials $a, b \in F[x]$ such that

$$\text{hcf}(f, g) = af + bg.$$

Proof We consider the set of polynomials

$$S = \{af + bg : a, b \in F[x]\}.$$

Compare the proof of this theorem to the proof of Proposition 4.7 in Book A, Chapter 1.

First note that if h is a polynomial in S, any associate of h is also in S, since these are formed by multiplying h by an element of F. By applying the Well-Ordering Principle to the degrees of the polynomials in S, we can choose from S a polynomial of smallest degree, $h = af + bg$. Moreover, since every associate of h is also in the set S, we can assume h is monic.

We claim that this monic polynomial h is unique. If there existed in S another monic polynomial $h_1 = a_1 f + b_1 g$ of the same degree as h, then

$$\begin{aligned} h - h_1 &= af + bg - (a_1 f + b_1 g) \\ &= (a - a_1)f + (b - b_1)g \end{aligned}$$

is another polynomial from S. Since h and h_1 are both monic, it follows that either $h - h_1 \neq 0_x$ has smaller degree than h (which is impossible), or $h - h_1 = 0_x$, so that h and h_1 are identical.

To complete the proof, we need to prove that $h = \text{hcf}(f, g)$. For this, we use the Division Algorithm (Theorem 3.1) to find polynomials q and r such that $f = qh + r$ with $\deg(r) < \deg(h)$. We want to show that $r = 0_x$. To do this, we demonstrate that r is a member of S:

$$\begin{aligned} r &= f - qh \\ &= 1 \cdot f - q\,(af + bg) \\ &= (1 - qa)f + (-qb)g \in S. \end{aligned}$$

However, r has strictly smaller degree than h, and so we must have $r = 0_x$. Hence $f = qh$ and so f is a multiple of h. A similar argument tells us that g is a multiple of h, and hence h is a common factor.

Finally, we have to show that h is the *highest* of these factors. This follows by Exercise 3.2(c): for any common factor d of f and g, we know that d divides $af + bg = h$. ■

Both Definition 3.4 and Theorem 3.5 have the condition that f and g cannot both be equal to the zero polynomial 0_x of $F[x]$. When working with the integers in Book A, Chapter 1 we simply stated that $\text{hcf}(0, 0)$ does not exist, and this would seem to make sense here too; thus, we declare that $\text{hcf}(0_x, 0_x)$ does not exist.

Some authors instead define $\text{hcf}(0_x, 0_x) = 0_x$.

Following the pattern laid out by Chapter 1, we can now give meaning to the term 'coprime' in the context of polynomials.

Definition 3.6 *Coprime*

Let F be a field, and let f, g be polynomials in $F[x]$, not both equal to 0_x. If $\text{hcf}(f, g) = 1_x$ then we say that f and g are **coprime**.

As a direct consequence of Theorem 3.5, we therefore have the following result.

Corollary 3.7

The polynomials $f, g \in F[x]$, not both equal to 0_x, are coprime if, and only if, there exist polynomials $a, b \in F[x]$ such that $af + bg = 1_x$.

Exercise 3.3

Let F be a field, and f, g and h be polynomials in $F[x]$. Prove the following.

(a) If $\text{hcf}(f, g) = 1_x$ and both f and g divide h, then fg divides h.

(b) If f divides gh and $\text{hcf}(f, g) = 1_x$, then f divides h.

Part (b) is the polynomial ring equivalent to Euclid's Lemma, Theorem 4.12 of Book A, Chapter 1.

3.3 The Euclidean Algorithm

In the last subsection, we proved the existence and uniqueness of the highest common factor of two polynomials over a field. However, as was the case in Book A, Chapter 1, the proof of this fact was not hugely enlightening on how to find the highest common factor. In this subsection, we will see how to apply the Division Algorithm (Theorem 3.1) to carry out practical calculation of the highest common factor.

Strategy: Euclidean Algorithm to find the hcf of two polynomials

Given a field F, and two polynomials $f, g \in F[x]$, not both equal to 0_x, let $f^* = f$ and $g^* = g$.

1. Apply the Division Algorithm (Theorem 3.1) to f^*, g^* to find $q, r \in F[x]$ for which $f^* = qg^* + r$.
2. If $r = 0_x$ then stop: $\text{hcf}(f, g) = a^{-1}g^*$, where $a \in F$ is the coefficient of the highest power of x in g^*.
3. Otherwise, $r \neq 0_x$. Replace f^* by g^*, and g^* by r, and go back to step 1.

Worked Exercise 3.8

In $\mathbb{Q}[x]$, find $\text{hcf}(3x^4 + 2x^3 + x^2 - 4x + 1,\ x^2 + x + 1)$.

Solution

For convenience later, we will set $f(x) = 3x^4 + 2x^3 + x^2 - 4x + 1$ and $g(x) = x^2 + x + 1$, so that we are trying to find $\text{hcf}(f, g)$. First, we divide f by g using polynomial long division, to obtain

$$f(x) = (3x^2 - x - 1)g(x) + (-2x + 2).$$

The next step of the Euclidean Algorithm requires that we divide $g(x)$ by the remainder, $-2x + 2$. This gives

$$g(x) = (-\tfrac{1}{2}x - 1)(-2x + 2) + 3.$$

As has happened here, when the remainder is a constant (that is, has degree zero) we know that the next step of the algorithm must give us zero remainder, and so the following final step could actually be omitted:

$$-2x + 2 = (-\tfrac{2}{3}x + \tfrac{2}{3}) \cdot 3 + 0.$$

Therefore, $\text{hcf}(f, g) = \frac{1}{3} \cdot 3 = 1$.

Exercise 3.4

(a) Use the Euclidean Algorithm to find $\text{hcf}(x^3 + 2x^2 - x - 2, x^2 - 4x + 3)$ in $\mathbb{Q}[x]$.

(b) Hence, or otherwise, find polynomials s, t in $\mathbb{Q}[x]$ for which $x - 1 = s(x^3 + 2x^2 - x - 2) + t(x^2 - 4x + 3)$.

As well as providing a method to find the hcf of two polynomials, the above exercise shows that the Euclidean Algorithm also helps us to find the polynomials $a(x), b(x)$ such that $\text{hcf}(f, g) = af + bg$ (as was guaranteed by Theorem 3.5). This was a relatively straightforward task in the above exercise, but for more complicated scenarios the strategy is to work backwards through the Euclidean Algorithm (as we did back in Book A, Chapter 1). We demonstrate this in a worked exercise.

Worked Exercise 3.9

Let $f(x) = 3x^4 + 2x^3 + x^2 - 4x + 1$ and $g(x) = x^2 + x + 1$. In $\mathbb{Q}[x]$, find polynomials $a(x)$ and $b(x)$ such that $1 = a(x)f(x) + b(x)g(x)$.

Solution

In Worked Exercise 3.8, we showed that $\text{hcf}(f, g) = 1$. Working backwards through the Euclidean Algorithm:

$$\begin{aligned} 3 &= (\tfrac{1}{2}x + 1)(-2x + 2) + g(x) \\ &= (\tfrac{1}{2}x + 1)\left[f(x) - (3x^2 - x - 1)g(x)\right] + g(x) \\ &= (\tfrac{1}{2}x + 1)f(x) - (\tfrac{3}{2}x^3 + \tfrac{5}{2}x^2 - \tfrac{3}{2}x - 2)g(x). \end{aligned}$$

Therefore, dividing by 3 we obtain $a(x) = \frac{1}{6}x + \frac{1}{3}$ and $b(x) = -\frac{1}{2}x^3 - \frac{5}{6}x^2 + \frac{1}{2}x + \frac{2}{3}$.

Once you have found polynomials $a(x)$ and $b(x)$, it is good practice to check they are correct by computing $af + bg$ directly.

Exercise 3.5

Let $f(x) = 2x^3 + 3x^2 + 2x - 1$ and $g(x) = x^2 + x$ be polynomials in $\mathbb{Q}[x]$.

(a) Find the highest common factor of f and g.

(b) Find polynomials $a(x)$ and $b(x)$ such that $af + bg = \text{hcf}(f, g)$.

3.4 Least common multiples

Throughout this section, we have been drawing parallels between the integers and polynomials with coefficients from fields. Here, our aim is to convert the results of Subsection 4.3 of Book A, Chapter 1.

Definition 3.10 *Least common multiple of two polynomials*

Let F be a field, and f, g two non-zero polynomials in $F[x]$. Then the **least common multiple** of f and g, $\operatorname{lcm}(f, g)$, is the monic polynomial $\ell \in F[x]$ satisfying the following:

(a) f and g both divide ℓ.

(b) For any polynomial $h \in F[x]$ that is divisible by both f and g, we have $\deg(h) \geq \deg(\ell)$.

First, we should establish whether $\operatorname{lcm}(f, g)$ must always exist. This is indeed the case because for any two non-zero polynomials f and g, the product fg and its associates satisfy condition (a) of Definition 3.10, and so there is a monic polynomial that satisfies condition (a). Among the monic polynomials satisfying (a), we need to choose one of smallest degree. Insisting that the least common multiple of two polynomials be monic then guarantees uniqueness, as we are about to show.

Lemma 3.11

Let F be a field, and f, g two non-zero polynomials in $F[x]$. Then $\ell = \operatorname{lcm}(f, g)$ is unique.

The proof follows the same pattern as the uniqueness part of the proof of Theorem 3.5.

Proof By the remarks before the lemma, it suffices to show that among the monic polynomials satisfying condition (a) of Definition 3.10, there is only one of lowest degree.

For a contradiction, suppose that ℓ_1 and ℓ_2 are two monic polynomials of the same lowest degree k that satisfy condition (a). However, then $\ell_1 - \ell_2$ is a polynomial of degree strictly less than k, and by Exercise 3.2(c), since each of f and g divides both ℓ_1 and ℓ_2, each also divides $\ell_1 - \ell_2$. Thus $\ell_1 - \ell_2$ is a polynomial of degree less than k that satisfies condition (a), which is a contradiction. ■

Now we have established that lcms are indeed unique, we want a practical way to find them. Of course, if we can factorise the polynomials f and g then it is relatively straightforward to find $\text{lcm}(f, g)$ (and indeed $\text{hcf}(f, g)$), simply by comparing factors as we would do with prime factorisations in the integers.

There are two problems with this approach. First, we haven't yet developed an analogue to 'prime factorisation' for polynomials (we will consider this task in the next section). Second, having managed to describe an analogue to 'prime factorisation', we would then need to find a way of calculating it. This is no easy task for integers, let alone polynomials over arbitrary fields! Instead, we will use the following result.

Proposition 3.12

Let $f(x), g(x)$ be non-zero monic polynomials with coefficients from a field F. Then

$$\text{lcm}(f, g) \cdot \text{hcf}(f, g) = f \cdot g.$$

Notice that this requires both f and g to be monic, otherwise fg is a scalar multiple of the left-hand side. Given non-monic polynomials f and g, we can obtain monic ones with the same lcm and hcf by multiplying each of f and g by the inverse of its leading coefficient.

Compare the proof of this proposition with the proof of Proposition 4.15 in Book A, Chapter 1: it is almost identical.

Proof Let $h(x) = \text{hcf}(f, g)$. Since h divides f, it also divides fg. Therefore, there exists a polynomial $\ell(x) \in F[x]$ such that $fg = h\ell$. We want to show that $\ell = \text{lcm}(f, g)$.

Since h divides both f and g, there exist polynomials p and q in $F[x]$ such that

$$f = hp \quad \text{and} \quad g = hq.$$

Remember that Lemma 2.10 of this chapter tells us that we are allowed to cancel factors from either side of an equation.

Substituting for f and then g in $fg = h\ell$, and then cancelling out a factor of h, we obtain

$$\ell = gp \quad \text{and} \quad \ell = fq.$$

This shows that ℓ is a common multiple of f and g, and so all that remains is for us to show it is the least such.

Suppose ℓ_1 is another polynomial in $F[x]$ that is divisible by both f and g. There exist polynomials $p_1, q_1 \in F[x]$ such that

$$\ell_1 = q_1 f \quad \text{and} \quad \ell_1 = p_1 g.$$

We also know from Theorem 3.5 that there exist polynomials a and b in $F[x]$ such that $h = \text{hcf}(f, g) = af + bg$.

Multiplying this equation by ℓ_1 gives

$$\begin{aligned}\ell_1 h &= \ell_1 af + \ell_1 bg \\ &= p_1 gaf + q_1 fbg \\ &= fg(p_1 a + q_1 b) \\ &= h\ell(p_1 a + q_1 b).\end{aligned}$$

This last line follows since we have already established that $fg = h\ell$. Finally, we can cancel a factor of h from either side to obtain

$$\ell_1 = \ell(p_1 a + q_1 b),$$

which proves that ℓ_1 is divisible by ℓ, and hence that $\ell(x) = \operatorname{lcm}(f, g)$. ■

Worked Exercise 3.13

Find $\operatorname{lcm}(x^3 + x^2 - x + 2, x^4 + 2x^3 - x - 2)$ in $\mathbb{Q}[x]$.

Solution

First, we compute $\operatorname{hcf}(x^3 + x^2 - x + 2, x^4 + 2x^3 - x - 2)$. The first round of polynomial long division tells us that

$$x^4 + 2x^3 - x - 2 = (x + 1)(x^3 + x^2 - x + 2) + (-2x - 4).$$

The remainder here is $-2x - 4 = -2(x + 2)$. Dividing $x^3 + x^2 - x + 2$ by the monic polynomial $x + 2$ gives

$$x^3 + x^2 - x + 2 = (x^2 - x + 1)(x + 2)$$

so that the hcf is $x + 2$.

Now, using Proposition 3.12,

$$\begin{aligned}&\operatorname{lcm}(x^3 + x^2 - x + 2, x^4 + 2x^3 - x - 2) \\ &\qquad = \frac{(x^3 + x^2 - x + 2) \cdot (x^4 + 2x^3 - x - 2)}{x + 2} \\ &\qquad = \frac{x^3 + x^2 - x + 2}{x + 2} \cdot (x^4 + 2x^3 - x - 2) \\ &\qquad = (x^2 - x + 1) \cdot (x^4 + 2x^3 - x - 2) \\ &\qquad = x^6 + x^5 - x^4 + x^3 - x^2 + x - 2.\end{aligned}$$

Exercise 3.6

Working in $\mathbb{Q}[x]$, find the lcm of:

(a) $x^3 - 1$ and $x^4 - x^3 + x^2 - 1$

(b) $x^3 + x^2 - 8x - 12$ and $x^3 + 5x^2 + 8x + 4$.

4 Factorising polynomials

This section explores methods of factorising polynomials into their 'indivisible' components.

4.1 Roots and irreducible polynomials

In the previous section, we learnt how to divide polynomials where the coefficients come from an arbitrary field, and this enabled us to compute hcfs and lcms. Another common task is to find roots: you will be familiar with exercises that ask for the values of $x \in \mathbb{R}$ for which $f(x) = 0$, and we will begin by generalising this to arbitrary fields.

Definition 4.1 *Root*

Let F be a field and let $f(x)$ be a non-zero polynomial in $F[x]$. Then $a \in F$ is called a **root** of $f(x)$ if $f(a) = 0$.

Notice that the above definition requires that the root a comes from the field F, that is, from the same field as the coefficients of the polynomial f. Why is this important? Consider, for example, the polynomial $f(x) = x^2 - 2$. As a polynomial in $\mathbb{R}[x]$, we can write $f(x) = (x - \sqrt{2})(x + \sqrt{2})$. On the other hand, as a polynomial in $\mathbb{Q}[x]$ we can see that $f(x)$ has no roots. We will explore this a little further in the next subsection, but note for now that the field over which a polynomial is defined (even if the coefficients come from a smaller field) really does matter.

Our first result should come as no surprise. It says that if we can find a root of a polynomial, then we can find a corresponding linear factor, and vice versa.

Theorem 4.2 *Factor Theorem*

Let F be a field, $a \in F$ and $f \in F[x]$ with $f \neq 0_x$. Then a is a root of f if, and only if, $f(x) = (x - a)q(x)$ for some $q(x) \in F[x]$, with $\deg(q) = \deg(f) - 1$.

Proof If $f(x) = (x - a)q(x)$, then clearly $f(a) = 0$.

For the converse, suppose that $f(a) = 0$. By the Division Algorithm (Theorem 3.1), there exist polynomials $q(x)$ and $r(x)$ such that $f(x) = (x - a)q(x) + r(x)$, where $\deg(r) < \deg(x - a) = 1$. Since $\deg(r) < 1$, $r(x)$ is a constant polynomial, and since $r(a) = 0$ it follows in fact that this constant must be zero, that is, $r(x) = 0_x$. Thus $f(x) = (x - a)q(x)$ as required. Finally, since $\deg(x - a) = 1$, we use Lemma 2.6 to obtain $\deg(f) = 1 + \deg(q)$. ■

We can deduce from the Factor Theorem that a polynomial in $F[x]$ of degree n has at most n roots. Loosely speaking, if a polynomial has more than n roots then it would have more than n factors, which if multiplied together would give a polynomial whose degree was greater than n.

Proposition 4.3

Let $f(x)$ be a non-zero polynomial of degree n in $F[x]$. Then the equation $f(x) = 0$ has at most n distinct roots in F.

Note that Lagrange's Theorem (Theorem 4.3 of Book A, Chapter 4) proves this result for the fields $\mathbb{Z}_p$, where p is prime.

Proof We will use induction on n, the degree of the polynomial. When $n = 1$, the polynomial is linear, and so we can write $f(x) = ax + b$. This has exactly one root as required.

Therefore, we will suppose $n \geq 2$, and suppose by induction that all polynomials of degree $m < n$ have at most m roots. If our polynomial $f(x)$ has no roots in F, then there is nothing more to prove, and so we may suppose that $a \in F$ is a root of $f(x)$. By the Factor Theorem 4.2, we have $f(x) = (x - a)g(x)$ for some polynomial $g(x) \in F[x]$ with $\deg(g) = n - 1$. By induction, $g(x)$ has at most $n - 1$ distinct roots in F.

Next, if $b \neq a$ is another root of $f(x)$, then it is also a root of $g(x)$. This follows because F has no zero divisors, and so, since $(b - a) \neq 0$, the equation $0 = f(b) = (b - a) \cdot g(b)$ implies that $g(b) = 0$. Thus, f cannot have more than n roots because otherwise g would have more than $n - 1$ roots, which is impossible. ■

Worked Exercise 4.4

Working in $\mathbb{Q}[x]$, factorise the following polynomials.

(a) $f(x) = x^3 + 3x^2 - 4$

(b) $f(x) = x^3 + 2x^2 + x$

Solution

(a) Note that $f(1) = 0$ and so $x - 1$ is a factor. Thus (using polynomial long division)

$$x^3 + 3x^2 - 4 = (x - 1)(x^2 + 4x + 4).$$

We can now factorise the remaining quadratic, either by inspection or using the quadratic equation, to get $f(x) = (x - 1)(x + 2)^2$.

(b) We can see immediately that x is a factor. Moreover, $f(-1) = 0$ and so $x + 1$ is a factor. We have $x^3 + 2x^2 + x = x(x + 1)(x + 1)$.

Exercise 4.1

Factorise the following polynomials over the stated field.

(a) $f(x) = x^4 - 3x^2 + 2$ over $\mathbb{R}$

(b) $f(x) = x^4 - 3x^2 + 2$ over $\mathbb{Q}$

(c) $f(x) = x^3 + x^2 + 2x + 2$ over $\mathbb{Z}_5$

(d) $f(x) = x^3 + x^2 + 2x + 2$ over $\mathbb{C}$

(e) $f(x) = x^4 - 6x^3 + 12x^2 - 10x + 3$ over $\mathbb{Q}$

Proposition 4.3 told us that a polynomial of degree n has at most n distinct roots, and we have already seen some examples where a polynomial has strictly fewer roots. As another example, consider the polynomial $x^2 + 1$: over $\mathbb{R}$ it has no roots, even though it is of degree 2. However, if we look at the same polynomial over $\mathbb{C}$ then we can find two roots, namely i and $-i$, and so

$$x^2 + 1 = (x + i)(x - i).$$

In fact, over the complex numbers a polynomial of degree n *always* has precisely n roots, provided we count repeated roots. This result is known as the Fundamental Theorem of Algebra and, ironically, is a theorem in complex analysis. We will not prove it here.

If we find, as we did in the case of $x^2 + 1$ over the reals, that for a given polynomial we do not identify as many roots as we might have expected, the question is whether we can nevertheless factorise the polynomial into smaller terms. For example, over $\mathbb{R}$ we can write the polynomial $x^4 - 1$ as a product

$$x^4 - 1 = (x + 1)(x - 1)(x^2 + 1),$$

and by what we saw a moment ago we know that $x^2 + 1$ does not factorise any further. In a sense, terms such as $x^2 + 1$ over $\mathbb{R}$ behave as the primes do for the integers, and this motivates our next definition.

Definition 4.5 *Irreducibility*

Let F be a field, and $f \in F[x]$ a non-zero polynomial over F. We say that f is **irreducible** if it is not a unit and it has no proper factors.

We say that f is **reducible** if it is not a unit but it does have proper factors.

For example, the polynomial $x^2 + 1$ is irreducible in $\mathbb{R}[x]$. However, since $x^2 + 1 = (x + i)(x - i)$, it is *not* irreducible in $\mathbb{C}[x]$. As we have said before, it is important to remember which polynomial ring we are working in!

In reading the above definition, we should recall *exactly* what we mean by 'unit' and 'proper factor' in a polynomial ring $F[x]$ over a field. First, the

units of $F[x]$ are, by Lemma 2.9 of this chapter, simply the non-zero elements of F. Next, if $f \neq 0_x$ can be written as $f = gh$ where g, h are 'proper factors', then neither g nor h is a unit; that is, the degrees of g and h must each be at least 1.

By virtue of the Fundamental Theorem of Algebra, the only irreducible polynomials in $\mathbb{C}[x]$ have degree 1. What about over $\mathbb{R}[x]$? The Fundamental Theorem of Algebra tells us something here too: suppose f is an irreducible polynomial in $\mathbb{R}[x]$ of degree greater than 1. As a polynomial in $\mathbb{C}[x]$, we can factorise it completely into linear factors. Moreover, if $z = a + ib$ $(b \neq 0)$ is a complex root of f, then the complex conjugate $\bar{z} = a - ib$ of z is also a non-real root of $f(x)$ in $\mathbb{C}$, since $f = \bar{f}$ is a polynomial with coefficients in $\mathbb{R}$. Thus, we can 'pair up' the complex linear factors of $f(x)$:

$$(x - (a + ib))(x - (a - ib)) = x^2 - 2ax + a^2 + b^2.$$

This product is of degree 2 and has coefficients from $\mathbb{R}$. Since it divides f and f is irreducible in $\mathbb{R}[x]$, it must be an associate of f. Thus $\deg(f) = 2$.

You might therefore be wondering where the interest in irreducible polynomials lies! The situation becomes much more interesting over the field of rationals $\mathbb{Q}$, and it is these that we will be exploring for the rest of this section.

Exercise 4.2

(a) Show that any polynomial of degree 1 is irreducible in $\mathbb{Q}[x]$.

(b) Prove that $f(x) = ax^2 + bx + c$ is irreducible in $\mathbb{R}[x]$ if, and only if, $b^2 - 4ac < 0$.

Recall that the expression $b^2 - 4ac$ is the discriminant of the quadratic polynomial $f(x)$.

4.2 Polynomials in $\mathbb{Z}[x]$ and rational roots

For the remainder of this section, we will focus on establishing conditions so that we can say when a polynomial in $\mathbb{Q}[x]$ is irreducible. You may have noticed that almost all the examples of polynomials we have seen in this chapter actually have integer coefficients; that is, they belong to the set

$$\mathbb{Z}[x] = \{a_0 + a_1x + a_2x^2 + \cdots + a_nx^n : a_0, a_1, \ldots, a_n \in \mathbb{Z}, n \geq 0\}.$$

This is the **polynomial ring over** $\mathbb{Z}$. Checking that $\mathbb{Z}[x]$ is indeed a (commutative) ring follows the approach we used for $\mathbb{Q}[x]$, even though $\mathbb{Z}$ is not a field.

In addition to the fact that most of the polynomials we have encountered actually belong to $\mathbb{Z}[x]$, suppose that $f(x) = \sum_{i=0}^{n} a_ix^i$ is a polynomial in $\mathbb{Q}[x]$. This means that $a_i \in \mathbb{Q}$ for all $i = 0, 1, \ldots, n$, and so we can write $a_i = \dfrac{b_i}{c_i}$ where $b_i \in \mathbb{Z}$ and $c_i \in \mathbb{N}$ with $\text{hcf}(b_i, c_i) = 1$ for each i. If we multiply $f(x)$ by all of the denominators c_i, then we are left with a polynomial with integer coefficients. In other words, a polynomial in $\mathbb{Q}[x]$

has an associate that is a polynomial in $\mathbb{Z}[x]$. The following lemma ensures that the resulting polynomial has the same irreducibility properties as f.

Lemma 4.6

Let f, g be polynomials in $\mathbb{Q}[x]$, such that f and g are associates. Then f is irreducible if, and only if, g is irreducible.

We will prove this result via the contrapositive statement of the lemma: f is reducible if, and only if, g is reducible.

Proof Since f and g are associates, there exists $\frac{p}{q} \in \mathbb{Q}$ such that $g(x) = \frac{p}{q} f(x)$. Now, if $f(x)$ is reducible, then $f(x) = a(x)b(x)$ where a and b are polynomials in $\mathbb{Q}[x]$ each of degree at least 1. However, now we can write $g(x) = \left(\frac{p}{q} a(x)\right) \cdot b(x)$, which shows that $g(x)$ is also reducible. The reverse implication follows by a similar argument, after writing $f(x) = \frac{q}{p} g(x)$. ■

Therefore, we can now restrict our attention to polynomials with integer coefficients, while continuing to consider irreducibility over $\mathbb{Q}[x]$. We will begin with a simple test that will enable us to determine whether a polynomial in $\mathbb{Z}[x]$ has any roots in $\mathbb{Q}[x]$.

Theorem 4.7 *Rational Root Test*

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial in $\mathbb{Z}[x]$ of degree $n \geq 1$, with a_n and a_0 non-zero. Then any rational root $\frac{p}{q} \in \mathbb{Q}$ of $f(x)$ with $\text{hcf}(p, q) = 1$ must satisfy the following:

(a) p is an integer factor of a_0

(b) q is an integer factor of a_n.

Proof Since $f\left(\frac{p}{q}\right) = 0$, we have

$$a_n \left(\frac{p}{q}\right)^n + a_{n-1} \left(\frac{p}{q}\right)^{n-1} + \cdots + a_1 \left(\frac{p}{q}\right) + a_0 = 0.$$

Moving the constant term a_0 to the right-hand side, and multiplying the equation by q^n, we obtain

$$p\left(a_n p^{n-1} + a_{n-1} q p^{n-2} + \cdots + a_1 q^{n-1}\right) = -a_0 q^n.$$

On the left-hand side, the expression in brackets is an integer (since all the terms are integers) and therefore the left-hand side is a multiple of p. This implies that p must divide $a_0 q^n$, but since $\text{hcf}(p, q) = 1$, it follows that p divides a_0.

Similarly, instead of moving the constant term to the right-hand side, we can move the leading term over, to obtain

$$q\left(a_{n-1} p^{n-1} + a_{n-2} q p^{n-2} + \cdots + a_0 q^{n-1}\right) = -a_n p^n.$$

By the same argument as before, q must divide $a_n p^n$, but since $\text{hcf}(p, q) = 1$ we must have $q \mid a_n$. ■

This theorem provides us with a method to find all rational roots of a polynomial in $\mathbb{Z}[x]$, and therefore of any polynomial in $\mathbb{Q}[x]$.

Worked Exercise 4.8

Find all the rational roots of $f(x) = 2x^3 + 3x - 5$.

Solution

By Theorem 4.7, if $\frac{p}{q} \in \mathbb{Q}$ is a solution to $f(x) = 0$, then p must divide 5 and q must divide 2.

Therefore, the possible values for p are ± 1, ± 5, and the possible values for q are 1, 2. This implies that the list of values of $\frac{p}{q}$ that *could* be roots of $f(x)$ is as follows.

Notice that we need to include negative values of p too.

$$\begin{aligned}\frac{p}{q} &= \pm\tfrac{1}{1}, \pm\tfrac{1}{2}, \pm\tfrac{5}{1}, \pm\tfrac{5}{2} \\ &= \pm 1, \pm\tfrac{1}{2}, \pm 5, \pm\tfrac{5}{2}\end{aligned}$$

We now need to check each possibility in turn, to determine whether any of these values are roots:

$$\begin{array}{llll} f(1) = 0 & f(-1) = -10 & f(\tfrac{1}{2}) = -\tfrac{13}{4} & f(-\tfrac{1}{2}) = -\tfrac{27}{4} \\ f(5) = 260 & f(-5) = -270 & f(\tfrac{5}{2}) = \tfrac{135}{4} & f(-\tfrac{5}{2}) = -\tfrac{175}{4}. \end{array}$$

From this, we conclude that the only root of $f(x)$ in $\mathbb{Q}$ is $x = 1$.

Exercise 4.3

Find the roots, if any, in $\mathbb{Q}$ of:

(a) $4x^4 - 2x^3 - 2x^2 - x + 1$

(b) $x^4 - x^3 - 2x^2 - 2x + 4$.

Exercise 4.4

Is $f(x) = 2x^3 - x^2 + 4x + 1$ irreducible in $\mathbb{Q}[x]$?

4.3 Gauss's Lemma

Given a polynomial in $\mathbb{Z}[x]$, can we determine whether it is irreducible in $\mathbb{Q}[x]$ by looking at its irreducibility in $\mathbb{Z}[x]$? The main result of this subsection, Gauss's Lemma, will tell us that we can. However, first we need to say what we mean by 'irreducible' in $\mathbb{Z}[x]$.

Definition 4.9 *Irreducibility over $\mathbb{Z}$*

Let f, g be polynomials in $\mathbb{Z}[x]$, with $f \neq 0_x$.

(a) We say that g **divides** f (or is a **factor** of f) if there exists some $h \in \mathbb{Z}[x]$ such that $f = gh$.

(b) If $f = gh$ and neither g nor h is equal to ± 1, then g, h are **proper factors**.

(c) We say that f is **irreducible** over $\mathbb{Z}$ if $f \neq \pm 1$ and it has no proper factors in $\mathbb{Z}[x]$.

Recall that the units of $\mathbb{Z}$ are 1 and -1.

We have changed very little to extend our definitions to polynomials over $\mathbb{Z}$, but there is one important point that you need to be aware of.

Example 4.10 *Factorisation in different polynomial rings*

Consider the polynomial $f(x) = 2x^2 + 2$.

(a) In $\mathbb{Z}[x]$, the factorisation $f(x) = 2 \cdot (x^2 + 1)$ is a proper factorisation, since neither factor is equal to ± 1, which are the units of $\mathbb{Z}$.

(b) In $\mathbb{Q}[x]$, the factorisation $f(x) = 2 \cdot (x^2 + 1)$ is *not* proper, since 2 is a unit. In fact, $f(x)$ is an associate of $x^2 + 1$, which is irreducible in $\mathbb{Q}[x]$ by Lemma 4.6.

At first sight, the above example might suggest that irreducibility in $\mathbb{Q}[x]$ is very different from irreducibility in $\mathbb{Z}[x]$. However, the following two definitions will enable us to proceed.

Definition 4.11 *Content of a polynomial*

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial in $\mathbb{Z}[x]$.

(a) The **content** of f is the highest common factor of all its coefficients:

$$c_f = \text{hcf}(a_0, a_1, \ldots, a_n).$$

(b) $f(x)$ is **primitive** if $c_f = 1$, or, in other words, if its coefficients have no non-trivial common factor.

For example, the content of the polynomial $6x^3 + 9x^2 - 3x + 18$ is $\text{hcf}(6, 9, -3, 18) = 3$.

We can eliminate the discrepancy encountered in Example 4.10 by restricting our view to primitive polynomials in $\mathbb{Z}[x]$. Note that if f is a polynomial in $\mathbb{Z}[x]$, then $f(x) = c_f f^*(x)$ for some primitive polynomial $f^*(x) \in \mathbb{Z}[x]$. Viewing f and f^* as polynomials over $\mathbb{Q}$, we can see that they are associates since the content c_f is an integer, and therefore a rational number. By Lemma 4.6, f^* is irreducible in $\mathbb{Q}[x]$ if, and only if, f is. So, from now on it suffices to prove results relating to irreducibility

using primitive polynomials only, and then the extension to all polynomials follows directly.

We require one last result before we state and prove Gauss's Lemma. The following statement looks innocent enough, but you may be surprised by how much work is required to prove it!

Lemma 4.12

The polynomials f and g are primitive in $\mathbb{Z}[x]$ if, and only if, their product fg is a primitive polynomial in $\mathbb{Z}[x]$.

Proof First we prove that if fg is primitive then f and g must also be primitive. Heading for a contradiction, suppose (without loss of generality) that f is not primitive. Then by writing $f = c_f f^*$, where c_f is the (non-unit) content of f and f^* is a primitive polynomial in $\mathbb{Z}[x]$, we see that $fg = (c_f f^*)g = c_f(f^*g)$ is also not primitive, proving one direction of the lemma.

Conversely, suppose that $f(x) = \sum_{i=0}^{n} a_i x^i$ and $g(x) = \sum_{i=0}^{m} b_i x^i$ are two primitive polynomials in $\mathbb{Z}[x]$, but that their product fg is not primitive. Therefore there exists a prime number $p \in \mathbb{Z}$ that divides the content of fg.

Since f and g are primitive, we know that p does not divide all the coefficients of f or of g. Therefore, we can find coefficients a_r of f and b_s of g such that p does not divide a_r or b_s. We pick r and s as large as possible, so that every coefficient of a higher-power term in each of f and g is divisible by p. Now consider the coefficient of x^{r+s} in fg:

$$\sum_{i+j=r+s} a_i b_j = a_0 b_{r+s} + a_1 b_{r+s-1} + \cdots + a_r b_s + \cdots + a_{r+s} b_0.$$

In this sum, when $i = r$ and $j = s$ then $a_i b_j$ is not divisible by p. However, for every other term $a_i b_j$ we have either $i > r$ or $j > s$, and so it must be divisible by p. Consequently, the coefficient of x^{r+s} cannot be divisible by p, which contradicts the fact that p divides the content of fg. Therefore, fg must be primitive. ■

Consider a primitive polynomial $f \in \mathbb{Z}[x]$: if $f = gh$, where $g, h \in \mathbb{Z}[x]$ are proper factors of f, then g and h are also proper factors of f in $\mathbb{Q}[x]$. Therefore, if f is reducible in $\mathbb{Z}[x]$, it is also reducible in $\mathbb{Q}[x]$. Gauss's Lemma proves the converse to this.

Theorem 4.13 *Gauss's Lemma*

Let $f(x)$ be a primitive polynomial in $\mathbb{Z}[x]$. If we can write $f = gh$ with $g, h \in \mathbb{Q}[x]$, then we can write $f = GH$ with $G, H \in \mathbb{Z}[x]$, $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Importantly, note that $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$, so that the proper factorisations of f in $\mathbb{Q}[x]$ and $\mathbb{Z}[x]$ are going to be closely connected.

Proof Since $g(x)$ is a polynomial in $\mathbb{Q}[x]$ we can find the lowest common denominator, b say, of all its coefficients. Hence we can write $g(x) = \frac{a}{b}G(x)$, where $G(x)$ is a primitive polynomial in $\mathbb{Z}[x]$ and $\frac{a}{b}$ is a fraction in its lowest terms. Similarly we can write $h(x) = \frac{c}{d}H(x)$, where $H(x)$ is a primitive polynomial in $\mathbb{Z}[x]$ and $\frac{c}{d}$ is a fraction in its lowest terms. Note that $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Hence $f(x) = g(x)h(x) = \frac{ac}{bd}G(x)H(x)$: we want to prove that $f = GH$, and so we need to show that $ac = bd$. Rearranging, we find $bdf = acGH$. Since f is primitive, the content of $bdf(x)$ is bd. Moreover, by Lemma 4.12 GH is primitive, and so the content of $acG(x)H(x)$ is ac. Since the polynomials $bdf(x)$ and $acG(x)H(x)$ are equal, so must their contents be equal, and hence $ac = bd$, as required. ■

Worked Exercise 4.14

Show that $f(x) = x^4 + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Solution

Note that if $f(x)$ has a factor of degree 3, it also has a factor of degree 1.

We need to show that $f(x)$ has no factors of degrees 1 or 2.

First we will eliminate the possibility of linear factors, using the Factor Theorem 4.2 and the Rational Root Test, Theorem 4.7. The latter tells us that any root $\frac{p}{q} \in \mathbb{Q}$ of $f(x)$ must have $p \mid 1$ and $q \mid 1$. Thus $p = \pm 1$, $q = 1$, and so the only numbers we need to check are ± 1. It is easy to verify that neither is a root of $f(x)$, and hence $f(x)$ has no linear factors.

This leaves the possibility that $f(x)$ factorises into two quadratic terms. By Gauss's Lemma, if $f(x)$ is reducible in $\mathbb{Q}[x]$ then it is reducible in $\mathbb{Z}[x]$, and so we can assume

$$f(x) = (ax^2 + bx + c)(rx^2 + sx + t)$$

where $a, b, c, r, s, t \in \mathbb{Z}$.

Multiplying out, we have

$$x^4 + x + 1 = arx^4 + (as + br)x^3 + (at + bs + cr)x^2 + (bt + cs)x + ct.$$

We can now compare coefficients between the left-hand side and the right-hand side. First, the coefficient of x^4 tells us that $ar = 1$, which implies that $a = r = \pm 1$. By multiplying all of a, b, c, r, s, t by -1 if necessary, we can assume $a = r = 1$.

Next, we consider the constant term, which tells us that $ct = 1$, and so $c = t = \pm 1$. Substituting $a = r = 1$ and $c = t = \pm 1$ into the equation above gives

$$x^4 + x + 1 = x^4 + (s + b)x^3 + (\pm 1 + bs \pm 1)x^2 \pm (b + s)x + 1.$$

The coefficient of x^2 gives $0 = \pm 2 + bs$, and the coefficient of x^3 gives $0 = s + b$, and so $b = -s$. Eliminating s, we find $b^2 = \pm 2$.

Neither $b^2 = 2$ nor $b^2 = -2$ has a solution in $\mathbb{Z}$, and so there can be no factorisation of $f(x)$ into quadratics in $\mathbb{Z}[x]$. Therefore $f(x)$ is irreducible in $\mathbb{Z}[x]$, and hence irreducible in $\mathbb{Q}[x]$ by Gauss's Lemma.

Exercise 4.5

Is $f(x) = x^4 - x^2 + 1$ irreducible in $\mathbb{Q}[x]$?

4.4 Further tests for irreducibility

Consider the polynomial $2x^5 - 3x^3 + 9x + 6$. Is it irreducible in $\mathbb{Q}[x]$? To answer this, we could modify the technique used in Worked Exercise 4.14, but the next theorem gives us a much quicker method.

Theorem 4.15 *Eisenstein's Criterion*

Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial in $\mathbb{Z}[x]$. If there is a prime $p \in \mathbb{Z}$ such that

(a) p divides a_i for $i = 0, \ldots, n-1$, but p does not divide a_n, and

(b) p^2 does not divide a_0,

then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Proof First, we may assume that $f(x)$ is primitive in $\mathbb{Z}[x]$ by the comments after Definition 4.11. By Gauss's Lemma, if $f = gh$ in $\mathbb{Q}[x]$ then $f = GH$ for some $G, H \in \mathbb{Z}[x]$ such that $\deg(G) = \deg(g)$ and $\deg(H) = \deg(h)$. Suppose that $G(x) = \sum_{i=0}^{r} b_i x^i$ and $H(x) = \sum_{j=0}^{s} c_j x^j$. Then, comparing coefficients gives

$$a_0 = b_0 c_0$$
$$a_1 = b_0 c_1 + b_1 c_0$$
$$a_k = \sum_{i=0}^{k} b_i c_{k-i} = b_0 c_k + b_1 c_{k-1} + \cdots + b_{k-1} c_1 + b_k c_0,$$
for each $k \in \{0, \ldots, n\}$.

(We set $b_i = 0$ for $i = r+1, \ldots, n$ and $c_j = 0$ for $j = s+1, \ldots, n$.)

Since p divides a_0, p divides either b_0 or c_0. Since p^2 does not divide a_0, p will not divide both b_0 and c_0. Suppose, without loss of generality, that p divides b_0 and not c_0.

Next, since p divides a_1 and not c_0, the second equation above shows that p divides b_1.

Now suppose that b_k is the first coefficient in $G(x)$ that is not divisible by p. Since

$$a_k = b_0c_k + b_1c_{k-1} + \cdots + b_{k-1}c_1 + b_kc_0,$$

only the last term in this sum is not divisible by p, and therefore p does not divide a_k. This is possible only if $k = n$, which implies that $b_n \neq 0$ and therefore $\deg(G) = n$. It follows that $\deg(g) = n$ and so $f(x)$ is irreducible in $\mathbb{Q}[x]$. ■

Worked Exercise 4.16

Show that the polynomial $2x^5 - 3x^3 + 9x + 6$ is irreducible in $\mathbb{Q}[x]$.

Solution

We apply Eisenstein's Criterion: $p = 3$ is a prime number and divides all the coefficients except the leading coefficient. $3^2 = 9$ does not divide 6, the constant term. Thus the conditions of Eisenstein's Criterion are satisfied and the polynomial is irreducible in $\mathbb{Q}[x]$.

Exercise 4.6

(a) Show that $x^n - p$ is irreducible in $\mathbb{Q}[x]$ for any prime number p.

(b) Show that $f(x) = \frac{3}{5}x^5 + \frac{5}{3}x^4 - x^3 + \frac{10}{3}x^2 + 2x - \frac{1}{3}$ is irreducible in $\mathbb{Q}[x]$.

One important family of polynomials in number theory is the **cyclotomic polynomials**, defined by

$$\Phi_n(x) = x^{n-1} + x^{n-2} + \cdots + x + 1 = \frac{x^n - 1}{x - 1}.$$

Φ is the Greek letter 'capital phi'.

Notice that the roots of $\Phi_n(x)$ in $\mathbb{C}$ satisfy $x^n = 1$ but $x \neq 1$ and thus consist of all the nth roots of unity, except for 1 itself. We will be interested in the case where $n = p$, a prime number, and we will use Eisenstein's Criterion to show that $\Phi_p(x)$ is irreducible in $\mathbb{Q}[x]$. However, we cannot use Eisenstein directly because all the coefficients of $\Phi_p(x)$ are equal to 1. Instead, we need a trick.

Lemma 4.17

Let $f(x)$ be a polynomial in $\mathbb{Z}[x]$. Then $f(x)$ is irreducible in $\mathbb{Q}[x]$ if, and only if, $f(x-b)$ is irreducible in $\mathbb{Q}[x]$ for any $b \in \mathbb{Q}$.

Proof Suppose $f(x) = g(x)h(x)$ in $\mathbb{Q}[x]$, where $g(x)$ and $h(x)$ are proper factors. Then $f(x-b) = g(x-b)h(x-b)$ is a proper factorisation of $f(x-b)$, since $\deg(g(x-b)) = \deg(g(x))$ and $\deg(h(x-b)) = \deg(h(x))$.

The reverse direction follows similarly: if $f(x-b)$ is reducible, then $f(x-b) = g(x-b)h(x-b)$ for some proper factors g, h. Setting $y = x-b$ we obtain $f(y) = g(y)h(y)$. From this we can deduce that $f(y-(-b)) = g(y-(-b))h(y-(-b))$. But $y-(-b) = x$ and so $f(x) = g(x)h(x)$. ■

Worked Exercise 4.18

Let p be a prime number. Prove that the cyclotomic polynomial $\Phi_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible in $\mathbb{Q}[x]$.

Solution

First, recall that $\Phi_p(x) = \dfrac{x^p - 1}{x - 1}$ and therefore, by applying the binomial theorem,

$$\Phi_p(x+1) = \frac{(x+1)^p - 1}{(x+1) - 1} = \frac{x^p + \binom{p}{1}x^{p-1} + \binom{p}{2}x^{p-2} + \cdots + \binom{p}{p-1}x}{x}$$

$$= x^{p-1} + \binom{p}{1}x^{p-2} + \binom{p}{2}x^{p-3} + \cdots + \binom{p}{p-1}.$$

Now $p \mid \binom{p}{i}$ for $1 \leq i \leq p-1$, and p^2 does not divide $\binom{p}{p-1} = p$, and so $\Phi_p(x+1)$ satisfies Eisenstein's Criterion. Therefore $\Phi_p(x+1)$ is irreducible in $\mathbb{Q}[x]$, and thus by Lemma 4.17 so also is $\Phi_p(x)$.

The **binomial coefficient** is defined as $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$.

Exercise 4.7

Prove that the following polynomials are irreducible in $\mathbb{Q}[x]$.

(a) $x^4 + 1$

(b) $x^8 + 1$

We finish with one final method for determining irreducibility of polynomials in $\mathbb{Z}[x]$, and hence, by Gauss's Lemma, in $\mathbb{Q}[x]$. This technique involves reducing the coefficients of the polynomial modulo some prime p.

Lemma 4.19

Let f be a polynomial in $\mathbb{Z}[x]$ and let p be a prime that does not divide the leading coefficient of f. If f is irreducible in $\mathbb{Z}_p[x]$, then f is irreducible in $\mathbb{Z}[x]$.

Before we prove this, consider the following example, which proves that the reverse of this lemma is not true.

Example 4.20 *Irreducibility in different polynomial rings*

The polynomial $x^2 + 3$ is irreducible in $\mathbb{Z}[x]$, but in $\mathbb{Z}_2[x]$ we have

$$x^2 + 3 \equiv x^2 + 1 \equiv x^2 + 2x + 1 \equiv (x + 1)^2 \pmod 2.$$

So $x^2 + 3$ is reducible in $\mathbb{Z}_2[x]$.

Note, however, that $x^2 + 3$ is irreducible in $\mathbb{Z}_5[x]$ since any proper factor must be a linear factor, and the equation $x^2 \equiv -3 \equiv 2 \pmod 5$ has no solution since 2 is not a quadratic residue modulo 5.

Proof of Lemma 4.19 Working towards a contradiction, suppose that f is irreducible in $\mathbb{Z}_p[x]$, but that $f = gh$ is a proper factorisation of f in $\mathbb{Z}[x]$. Let g^* and h^* denote the polynomials obtained from g and h by reducing the coefficients modulo p. We then have $f(x) \equiv g^*(x)h^*(x) \pmod p$, and since the leading coefficient of f is not divisible by p, it follows that $\deg(g) = \deg(g^*)$ and $\deg(h) = \deg(h^*)$, and therefore f is reducible in $\mathbb{Z}_p[x]$, which is a contradiction. ■

To use this lemma, the skill is to find a prime p such that reducing modulo p gives a polynomial that can be shown to be irreducible in $\mathbb{Z}_p[x]$.

Worked Exercise 4.21

Show that $f(x) = x^4 + 15x^2 + 7$ is irreducible in $\mathbb{Q}[x]$.

Solution

Note first that if we reduce modulo 2, we get

$$f(x) \equiv x^4 + x^2 + 1 \equiv (x^2 + x + 1)(x^2 + x + 1) \pmod 2,$$

which is not irreducible. Similarly, reducing modulo 3 also does not work because

$$f(x) \equiv x^4 + 1 \equiv (x^2 + x - 1)(x^2 - x - 1) \pmod 3.$$

Next, we try reducing modulo 5 to get

$$f(x) \equiv x^4 + 2 \pmod 5,$$

which is not obviously reducible. First, we test whether any of the five elements of $\mathbb{Z}_5$ is a root of $f(x)$ in $\mathbb{Z}_5[x]$, and we find that none is:

$$f(0) \equiv 2, \quad f(1) \equiv 3, \quad f(2) \equiv 3, \quad f(3) \equiv 3, \quad f(4) \equiv 3.$$

This leaves the possibility that $f(x) \equiv g(x)h(x) \pmod 5$, where g and h are quadratic factors in $\mathbb{Z}_5[x]$. Since f is monic, we can assume without loss of generality that both g and h are monic. So then we can write

$$x^4 + 2 \equiv (x^2 + ax + b)(x^2 + cx + d)$$

for some $a, b, c, d \in \mathbb{Z}_5$. Comparing coefficients of x^3, x^2 and x^0 in turn, we get the following equations:

$$\begin{aligned} a + c &\equiv 0 \pmod 5 \\ ac + b + d &\equiv 0 \pmod 5 \\ bd &\equiv 2 \pmod 5. \end{aligned}$$

Using $c \equiv -a$ from the first equation in the second tells us that $b + d \equiv a^2$, and so $b + d \equiv 0, 1$ or 4 since these are the only quadratic residues modulo 5. It follows that $d \equiv -b, 1 - b$ or $4 - b$. Using these three possible values of d in the third equation, $bd \equiv 2$, gives

$$\begin{aligned} b^2 &\equiv -2 \pmod 5 \\ b(1 - b) &\equiv 2 \pmod 5 \\ b(4 - b) &\equiv 2 \pmod 5. \end{aligned}$$

We can exhaustively check that none of the values $b = 0, 1, 2, 3, 4$ satisfies any of these equations. This tells us that $x^4 + 2$ has no quadratic factor modulo 5, and therefore that $f(x)$ is irreducible modulo 5. Hence, by Lemma 4.19, $f(x)$ is irreducible in $\mathbb{Z}[x]$, and hence irreducible in $\mathbb{Q}[x]$ by Gauss's Lemma.

Exercise 4.8

Are the following polynomials irreducible in $\mathbb{Q}[x]$? If not, give their factorisations into irreducible polynomials.

(a) $x^3 + 7x^2 - 14x + 4$

(b) $x^4 + 4$

Exercise 4.9

(a) List the irreducible polynomials of degree 4 or less in $\mathbb{Z}_2[x]$.

(b) Hence, or otherwise, prove that the following polynomials are irreducible over $\mathbb{Q}$.

(i) $x^4 - 3x^3 + 12x^2 - 2x + 5$

(ii) $3x^4 - x^3 + 71x^2 - 9x + 1$

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should:

(a) Recall and be able to use the axioms that define a ring, and know the basic properties of rings arising from these axioms.

(b) Understand the definition of 'subring', and the statement, proof and uses of the subring criterion.

(c) Know the meaning of and be able to use the terms 'zero divisor' and 'unit'.

(d) Understand the terms 'field' and 'subfield', and recall basic properties of these.

(e) Know how to add and multiply polynomials over arbitrary fields, and be able to use this to define polynomial rings.

(f) Understand the term 'degree of a polynomial', and recall results relating the degree of polynomials to the addition and multiplication operations.

(g) Recall basic properties of polynomial rings, such as the non-existence of zero divisors, classification of units and cancellation of polynomials.

(h) Understand the terms 'associate' and 'monic polynomial'.

(i) Understand the statement and proof of the Division Algorithm for polynomials, and be able to apply polynomial long division in the ring $\mathbb{Q}[x]$.

(j) Understand the meaning of the highest common factor of two polynomials, the proof of the existence of the hcf, and the meaning of 'coprime' in the context of polynomials over fields.

(k) Be able to apply the Euclidean Algorithm to compute the hcf of two polynomials f and g in $\mathbb{Q}[x]$, and find polynomials a, b such that $\mathrm{hcf}(f, g) = af + bg$.

(l) Understand the meaning of the least common multiple of two polynomials, the proof of its uniqueness, and be able to compute lcms in the polynomial ring $\mathbb{Q}[x]$.

(m) Understand the terms 'root', 'irreducible' and 'reducible' polynomial, and be able to recall and use the Factor Theorem to factorise polynomials defined over arbitrary fields.

(n) Understand and be able to apply the Rational Root Test to find roots of polynomials over $\mathbb{Q}$.

(o) Understand the terms 'content' and 'primitive', and recall the statement and proof of Gauss's Lemma.

(p) Be able to apply Gauss's Lemma, Eisenstein's Criterion and its variants to test for irreducibility of polynomials in $\mathbb{Q}[x]$.

Solutions to exercises

Solution to Exercise 1.1

(a) $(2,1)+(4,5)=(6,6)$ and $(2,1)\cdot(4,5)=(8,5)$

(b) $(0,0)$

(c) $(-2,-1)$

(d) $(1,1)$

Solution to Exercise 1.2

For the additive axioms, we could use the fact that $(\mathbb{Z}_n,+)$ is an abelian group.

(a) We check each axiom in turn. R1, R2 and R5 follow by the basic properties of addition modulo n. For R3, the additive identity is 0, since $0+a\equiv a\pmod{n}$ for any $a\in\mathbb{Z}_n$, and then writing $a+(-a)\equiv 0\pmod{n}$ proves R4.

The multiplicative axioms follow similarly: R6 and R7 follow directly from the definition of multiplication modulo n. For R8, the multiplicative identity is 1, since $1\cdot a=a\cdot 1\equiv a\pmod{n}$.

Finally, R9 follows directly from the distributivity of addition and multiplication of the integers, and R10 (commutativity of multiplication) holds by the properties of multiplication modulo n. Hence $(\mathbb{Z}_n,+,\cdot)$ is a commutative ring.

(b) For $\mathbb{Z}[\sqrt{2}]$, let $a,b,c,d\in\mathbb{Z}$. We consider each axiom in turn.

For R1, we have

$$(a+b\sqrt{2})+(c+d\sqrt{2})=(a+c)+(b+d)\sqrt{2},$$

which is in $\mathbb{Z}(\sqrt{2})$ since $a+c,b+d\in\mathbb{Z}$.

Using the equation in R1 repeatedly together with the associativity of addition of integers, R2 follows directly. The additive identity is $0=0+0\sqrt{2}$, which proves R3, and then for R4, the inverse of $a+b\sqrt{2}$ is $(-a)+(-b)\sqrt{2}$. Commutativity follows from the equation in R1 with the commutativity of addition over the real numbers.

For the multiplicative axioms, closure (R6) is not immediately obvious, but it follows from

$$\begin{aligned}(a+b\sqrt{2})\cdot(c+d\sqrt{2}) &= ac+(bc+ad)\sqrt{2}+bd\sqrt{2}\cdot\sqrt{2}\\ &= (ac+2bd)+(bc+ad)\sqrt{2}.\end{aligned}$$

R7 and R10 are now immediate since multiplication of real numbers is associative and commutative. For R8, the multiplicative identity is $1=1+0\sqrt{2}$. Finally, R9 follows, once again, because multiplication is distributive over addition in $\mathbb{R}$.

Solution to Exercise 1.3

We use part (c) of the proposition twice, followed by part (b):

$$(-a)\cdot(-b)=-(a\cdot(-b))=-(-(a\cdot b))=a\cdot b.$$

Solution to Exercise 1.4

If $1 = 0$ and a is any element of the ring then, using Proposition 1.5(a),

$$a = a \cdot 1 = a \cdot 0 = 0$$

and hence $a = 0$. Consequently, the ring has only one element, which must be both the additive and the multiplicative identity.

Solution to Exercise 1.5

(a) We show that $\mathbb{Z}[\omega]$ is a subring of $\mathbb{C}$. Let $a, b, c, d, e, f \in \mathbb{Z}$. For SR1, we have

$$(a + b\omega + c\omega^2) - (d + e\omega + f\omega^2) = (a - d) + (b - e)\omega + (c - f)\omega^2,$$

which is in $\mathbb{Z}[\omega]$ since $(a - d)$, $(b - e)$ and $(c - f) \in \mathbb{Z}$.

For SR2, using $\omega^3 = 1$, we have

$$\begin{aligned}&(a + b\omega + c\omega^2) \cdot (d + e\omega + f\omega^2)\\ &\quad = (ad + bf + ce) + (ae + bd + cf)\omega + (af + be + cd)\omega^2 \in \mathbb{Z}[\omega].\end{aligned}$$

Finally, the multiplicative identity of $\mathbb{C}$ is 1, and in $\mathbb{Z}[\omega]$ we have $1 = 1 + 0\omega + 0\omega^2$. Hence SR3 holds.

Note that $\mathbb{Z}[\omega]$ is *not* a subring of $\mathbb{Z}[i]$.

(b) We will show that $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$. For a, b, c, d in $\mathbb{Q}$, we have

SR1: $(a + ib) - (c + id) = (a - c) + i(b - d) \in \mathbb{Q}[i]$.

SR2: $(a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc) \in \mathbb{Q}[i]$.

SR3: The multiplicative identity of $\mathbb{C}$ and $\mathbb{Q}[i]$ is $1 = 1 + 0i$.

Solution to Exercise 1.6

(a) $(3, 0) \cdot (0, 1) = (0, 0)$ gives an example of two zero divisors.

(b) The zero divisors are those elements (x, y) for which one of x and y is 0.

Solution to Exercise 1.7

(a) The units are 1, 5, 7 and 11. Each element is its own inverse – for example, $7 \times 7 \equiv 1 \pmod{12}$.

(b) The units are the pairs $(1, 1)$, $(1, -1)$, $(-1, 1)$ and $(-1, -1)$.

(c) The units are $1, -1, i$ and $-i$.

Solution to Exercise 1.8

(a) The zero divisors in $\mathbb{Z}_8$ are 2, 4 and 6, since $2 \times 4 \equiv 6 \times 4 \equiv 0 \,(\text{mod } 8)$.

In $\mathbb{Z}_9$ the zero divisors are 3 and 6.

In $\mathbb{Z}_{10}$ the zero divisors are 2, 4, 5, 6, 8.

In $\mathbb{Z}_{11}$ there are no zero divisors.

In $\mathbb{Z}_{12}$ the zero divisors are 2, 3, 4, 6, 8, 9 and 10.

We will use this result in the proof of Lemma 1.18.

(b) The zero divisors in $\mathbb{Z}_n$ are those positive integers $a \in \mathbb{Z}_n$ that share a common factor with n that is greater than 1; that is, $\text{hcf}(a, n) > 1$.

(c) $\mathbb{Z}_n$ has no zero divisors if, and only if, n is a prime number.

(d) Suppose that $a \in R$ is a unit, and so there exists a multiplicative inverse a^{-1}. If there exists $b \in R$ such that $a \cdot b = 0$, then

$$b = 1 \cdot b = (a^{-1}a)b = a^{-1}(ab) = a^{-1} \cdot 0 = 0.$$

Therefore a cannot be a zero divisor.

Solution to Exercise 1.9

(a) By definition, $(a^{-1})^{-1} \cdot a^{-1} = 1$, and therefore we have

$$\begin{aligned}(a^{-1})^{-1} &= (a^{-1})^{-1} \cdot 1 \\ &= (a^{-1})^{-1} \cdot (a^{-1} \cdot a) \\ &= ((a^{-1})^{-1} \cdot a^{-1}) \cdot a \\ &= 1 \cdot a = a.\end{aligned}$$

(b)
$$\begin{aligned}b^{-1}a^{-1} &= (b^{-1}a^{-1}) \cdot 1 = (b^{-1}a^{-1}) \cdot \big((ab)(ab)^{-1}\big) \\ &= \big((b^{-1}a^{-1})(ab)\big) \cdot (ab)^{-1} \\ &= \big(b^{-1}(a^{-1}a)b\big) \cdot (ab)^{-1} = \big(b^{-1}b\big) \cdot (ab)^{-1} \\ &= 1 \cdot (ab)^{-1} = (ab)^{-1}.\end{aligned}$$

Alternatively, consider

$$\begin{aligned}(b^{-1}a^{-1}) \cdot (ab) &= b^{-1}(a^{-1}a)b \\ &= b^{-1} \cdot 1 \cdot b = b^{-1}b \\ &= 1,\end{aligned}$$

which shows that $(b^{-1}a^{-1})$ is the inverse of the element ab. That is, $(ab)^{-1} = b^{-1}a^{-1}$.

Solution to Exercise 1.10

(a) From Exercise 1.5(b), we know that $\mathbb{Q}[i]$ is a subring of $\mathbb{C}$. To show that $\mathbb{Q}[i]$ is a subfield of $\mathbb{C}$, it remains to show only that every non-zero element of $\mathbb{Q}[i]$ has a multiplicative inverse, by Lemma 1.21. Let $a, b \in \mathbb{Q}$, not both equal to zero, so that $a + bi \in \mathbb{Q}[i]$ is non-zero.

We want to find $c, d \in \mathbb{Q}$ such that $(a+bi)(c+di) = 1$. We could compare real and imaginary parts, and solve the resulting simultaneous equations, but there is a quicker method. Observe that $(a+bi)(a-bi) = a^2 + b^2$ is a rational number, and therefore

$$\frac{(a+bi)(a-bi)}{a^2+b^2} = 1.$$

Thus $c = \dfrac{a}{a^2+b^2}$ and $d = \dfrac{-b}{a^2+b^2}$ are both in $\mathbb{Q}$, and the multiplicative inverse $c + di \in \mathbb{Q}[i]$ is

$$\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i.$$

(b) We will use the fact that $\mathbb{Q}[\sqrt{2}] \subset \mathbb{R}$. First we show, using the subring criterion (Lemma 1.8) that $\mathbb{Q}[\sqrt{2}]$ is a subring of $\mathbb{R}$. For $a, b, c, d \in \mathbb{Q}$, we have

SR1: $(a + b\sqrt{2}) - (c + d\sqrt{2}) = (a-c) + (b-d)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$.

SR2: $(a + b\sqrt{2}) \cdot (c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$.

SR3: The multiplicative identity of $\mathbb{R}$ and $\mathbb{Q}[\sqrt{2}]$ is $1 = 1 + 0\sqrt{2}$.

Finally, we find the multiplicative inverses of elements in $\mathbb{Q}[\sqrt{2}]$. Let $a, b \in \mathbb{Q}$, with at least one of a or b non-zero, and consider $a + b\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$. We wish to find $c, d \in \mathbb{Q}$ such that $(a + b\sqrt{2})(c + d\sqrt{2}) = 1$.

Observe that $(a + b\sqrt{2})(a - b\sqrt{2}) = a^2 - 2b^2$ is in $\mathbb{Q}$, and is non-zero since $\sqrt{2}$ is irrational. Thus,

$$\frac{(a+b\sqrt{2})(a-b\sqrt{2})}{a^2-2b^2} = 1.$$

If $a^2 - 2b^2$ were 0 with at least one of a and b non-zero then b must be non-zero and so $\frac{a^2}{b^2} = 2$. This gives the contradiction that $\sqrt{2} = \frac{a}{b} \in \mathbb{Q}$.

Therefore, we may take

$$c = \frac{a}{a^2-2b^2}, \quad d = \frac{-b}{a^2-2b^2},$$

noting that this choice satisfies $c, d \in \mathbb{Q}$, and therefore $c + d\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$. Hence, the inverse of $a + b\sqrt{2}$ is

$$\frac{a}{a^2-2b^2} + \frac{b}{2b^2-a^2}\sqrt{2}.$$

Solution to Exercise 2.1

(a) Working in $\mathbb{Q}[x]$:

(i) $f(x) + g(x) = 9x^4 + 2x^3 + 7x^2 - x - 2$

(ii) $f(x)g(x) = 18x^7 + 14x^5 - 29x^4 + 2x^3 - 21x^2 + 3x - 3$.

(b) Working in $\mathbb{Z}_7[x]$, we can reduce the answer to part (a) modulo 7:

(i) $f(x) + g(x) \equiv 2x^4 + 2x^3 + 6x + 5 \pmod 7$

(ii) $f(x)g(x) \equiv 4x^7 + 6x^4 + 2x^3 + 3x + 4 \pmod 7$.

(c) Working in $\mathbb{Z}_2[x]$, we can reduce the answer to part (a) modulo 2:

(i) $f(x) + g(x) \equiv x^4 + x^2 + x \pmod 2$

(ii) $f(x)g(x) \equiv x^4 + x^2 + x + 1 \pmod 2$.

Solution to Exercise 2.2

(a) $\deg(f) = 2$

(b) $\deg(g) = 1$

(c) $$\begin{aligned} f(x)g(x) &= (2x^2 + x + 3)(3x + 1) \\ &= 2 \cdot 3x^3 + (2 + 3)x^2 + (3 \cdot 3 + 1)x + 3 \\ &\equiv 5x^2 + 4x + 3 \pmod 6. \end{aligned}$$

From this, we see that $\deg(fg) = 2$. Note that this is less than $\deg(f) + \deg(g) = 3$.

Solution to Exercise 2.3

(a) $f(x) + g(x) = (2x^3 + 4x^2 - 1) + (3x^3 + 3x^2 - 2x) = 5x^3 + 7x^2 - 2x - 1 \equiv 2x^2 - 2x - 1 \pmod 5$. Therefore $\deg(f + g) = 2$.

(b) $f(x) + g(x) = (1 + x) + (1 - x) = 2$, and therefore $\deg(f + g) = 0$.

Solution to Exercise 2.4

(a) Since $g(x)$ is an associate of $f(x)$, there exists a unit $u \in F[x]$ such that $g(x) = uf(x)$. By Lemma 2.6, $\deg(g) = \deg(uf)$ $= \deg(u) + \deg(f)$, and since $\deg(u) = 0$ by Lemma 2.9, we have $\deg(f) = \deg(g)$.

(b) Let $f(x)$ be a polynomial of degree $\deg(f) = n$ over F with leading coefficient $a_n \neq 0$. Since $a_n \in F$, there exists a multiplicative inverse $a_n^{-1} \in F$ of a_n, and so the polynomial $a_n^{-1}f(x)$ is an associate of $f(x)$, and has leading coefficient equal to 1. That is, the polynomial $a_n^{-1}f(x)$ is monic.

Moreover, it is the unique monic associate: a_n^{-1} is the unique inverse of a_n in F, and so any other associate of $f(x)$ cannot be monic.

Solution to Exercise 3.1

(a)

$$\begin{array}{r|llllll} & x^2 & - & x & + & 2 & \\ \hline x-1 & x^3 & - & 2x^2 & + & 3x & - \ 1 \\ & x^3 & - & x^2 & & & \\ \hline & & - & x^2 & + & 3x & - \ 1 \\ & & - & x^2 & + & x & \\ \hline & & & & & 2x & - \ 1 \\ & & & & & 2x & - \ 2 \\ \hline & & & & & & 1 \end{array}$$

Thus $q(x) = x^2 - x + 2$ and $r(x) = 1$.

(b)

$$\begin{array}{r|llllll} & 2x^2 & & - & 2 & & \\ \hline x^2+1 & 2x^4 & & & & - \ x & + \ 1 \\ & 2x^4 & & + & 2x^2 & & \\ \hline & & & - & 2x^2 & - \ x & + \ 1 \\ & & & - & 2x^2 & & - \ 2 \\ \hline & & & & & - \ x & + \ 3 \end{array}$$

Thus $q(x) = 2x^2 - 2$ and $r(x) = -x + 3$.

(c)

$$\begin{array}{r|llllll} & \frac{3}{2}x^2 & - & \frac{7}{4}x & + & \frac{7}{8} & \\ \hline 2x+1 & 3x^3 & - & 2x^2 & & & + \ 1 \\ & 3x^3 & + & \frac{3}{2}x^2 & & & \\ \hline & & - & \frac{7}{2}x^2 & & & + \ 1 \\ & & - & \frac{7}{2}x^2 & - & \frac{7}{4}x & \\ \hline & & & & & \frac{7}{4}x & + \ 1 \\ & & & & & \frac{7}{4}x & + \ \frac{7}{8} \\ \hline & & & & & & \frac{1}{8} \end{array}$$

Thus $q(x) = \frac{3}{2}x^2 - \frac{7}{4}x + \frac{7}{8}$ and $r(x) = \frac{1}{8}$.

Solution to Exercise 3.2

(a) Since g divides f, there exists a polynomial $a \in F[x]$ such that $f = ag$. Now, using the distributivity of addition and multiplication in $F[x]$, we have

$$f + gh = ag + gh = g(a + h),$$

which shows that g is a factor of $f + gh$.

(b) Since g divides f, we can write $f = ag$ for some $a \in F[x]$. Similarly, since h divides g, there exists $b \in F[x]$ such that $g = bh$. Now, by associativity of multiplication in $F[x]$,

$$f = ag = a(bh) = (ab)h,$$

which shows that f is divisible by h.

(c) Since h divides f and g, we can write $f = ch$ and $g = dh$ for some $c, d \in F[x]$. Now, using associativity of multiplication and the distributive laws, we see that h divides $af + bg$:

$$af + bg = a(ch) + b(dh) = (ac)h + (bd)h = (ac + bd)h.$$

(d) If f and g are associates then there exists $a \in F$ such that $g(x) = af(x)$, which immediately shows that g is divisible by f. Moreover, we can write $a^{-1}g(x) = f(x)$, which shows that f is divisible by g.

For the other direction, suppose g divides f and f divides g. Then there exist polynomials $a, b \in F[x]$ such that $g = af$ and $f = bg$. We will have finished if we can show that a and b are inverses, since this would imply $a \in F$, so that g is an associate of f. We write

$$f = bg = b(af) = (ba)f.$$

Then by Lemma 2.10 we have $1 = 1_x = ba$, and therefore $a = b^{-1}$ as required.

Solution to Exercise 3.3

(a) Since f and g are coprime, there exist $a, b \in F[x]$ such that $1_x = af + bg$. Multiplying this through by h gives

$$h = afh + bgh.$$

Next, as f divides h there exists $c \in F[x]$ such that $h = cf$. Similarly there exists $d \in F[x]$ such that $h = dg$, and we substitute these in to the right-hand side to get

$$h = af(dg) + bg(cf) = (ad + bc)fg,$$

which shows that fg divides h.

(b) Since $\mathrm{hcf}(f, g) = 1_x$, there exist $a, b \in F[x]$ such that $1_x = af + bg$. We multiply this through by h to get

$$h = afh + bgh.$$

Next, since f divides gh, there exists $c \in F[x]$ such that $gh = cf$. Substituting this into the above equation gives

$$h = afh + b(cf) = (ah + bc)f,$$

which shows that f divides h.

Solution to Exercise 3.4

(a) Using polynomial long division, we apply the Euclidean Algorithm:

$$(x^3 + 2x^2 - x - 2) = (x + 6)(x^2 - 4x + 3) + (20x - 20)$$
$$(x^2 - 4x + 3) = \tfrac{1}{20}(x - 3)(20x - 20) + 0.$$

Hence in $\mathbb{Q}[x]$,
$\mathrm{hcf}(x^3 + 2x^2 - x - 2, x^2 - 4x + 3) = \frac{1}{20}(20x - 20) = x - 1$.

(b) From the above calculation,
$x - 1 = \frac{1}{20}\left((x^3 + 2x^2 - x - 2) - (x + 6)(x^2 - 4x + 3)\right)$ and so
$x - 1 = \frac{1}{20}(x^3 + 2x^2 - x - 2) - \frac{1}{20}(x + 6)(x^2 - 4x + 3)$. That is,
$s = \frac{1}{20}$ and $t = -\frac{1}{20}(x + 6)$.

Solution to Exercise 3.5

(a) First, we divide $f(x) = 2x^3 + 3x^2 + 2x - 1$ by $g(x) = x^2 + x$ using polynomial long division:

$$f(x) = (2x + 1)g(x) + (x - 1).$$

Next, we divide $g(x)$ by the remainder, $x - 1$:

$$g(x) = (x + 2)(x - 1) + 2.$$

This tells us that $\mathrm{hcf}(f, g) = \frac{1}{2} \cdot 2 = 1$.

(b) Working backwards through the Euclidean Algorithm,

$$\begin{aligned} 2 &= g(x) - (x + 2)(x - 1) \\ &= g(x) - (x + 2)\left[f(x) - (2x + 1)g(x)\right] \\ &= -(x + 2)f(x) + (2x^2 + 5x + 3)g(x). \end{aligned}$$

Dividing by 2, this gives $a(x) = -\frac{1}{2}x - 1$, and $b(x) = x^2 + \frac{5}{2}x + \frac{3}{2}$.

Solution to Exercise 3.6

(a) First, polynomial long division gives

$$x^4 - x^3 + x^2 - 1 = (x - 1)(x^3 - 1) + (x^2 + x - 2).$$

Dividing $x^3 - 1$ by the remainder gives

$$x^3 - 1 = (x - 1)(x^2 + x - 2) + (3x - 3).$$

The remainder here is $3x - 3 = 3(x - 1)$ and we can see by inspection that $x^2 + x - 2 = (x - 1)(x + 2)$, which allows us to conclude that the hcf is $x - 1$. Therefore, by Proposition 3.12,

$$\begin{aligned} \mathrm{lcm}(x^3 - 1, x^4 - x^3 + x^2 - 1) &= \frac{(x^3 - 1) \cdot (x^4 - x^3 + x^2 - 1)}{x - 1} \\ &= (x^2 + x + 1) \cdot (x^4 - x^3 + x^2 - 1) \\ &= x^6 + x^4 - x - 1. \end{aligned}$$

(b) Dividing one by the other gives

$$x^3 + 5x^2 + 8x + 4 = 1 \cdot (x^3 + x^2 - 8x - 12) + (4x^2 + 16x + 16).$$

The remainder is $4x^2 + 16x + 16 = 4(x^2 + 4x + 4)$, and so we can now divide $x^3 + x^2 - 8x - 12$ by $x^2 + 4x + 4$ to get

$$x^3 + x^2 - 8x - 12 = (x - 3)(x^2 + 4x + 4).$$

Therefore $\text{hcf}(x^3 + x^2 - 8x - 12, x^3 + 5x^2 + 8x + 4) = x^2 + 4x + 4$.

Finally, using Proposition 3.12,

$$\begin{aligned}\text{lcm}(x^3 + x^2 - 8x - 12, x^3 + 5x^2 + 8x + 4) &= \frac{(x^3 + x^2 - 8x - 12) \cdot (x^3 + 5x^2 + 8x + 4)}{x^2 + 4x + 4} \\ &= (x - 3) \cdot (x^3 + 5x^2 + 8x + 4) \\ &= x^4 + 2x^3 - 7x^2 - 20x - 12.\end{aligned}$$

Alternatively, note that $x^3 + x^2 - 8x - 12 = (x - 3)(x + 2)^2$ and $x^3 + 5x^2 + 8x + 4 = (x + 1)(x + 2)^2$, and so

$$\text{lcm}(x^3 + x^2 - 8x - 12, x^3 + 5x^2 + 8x + 4) = (x - 3)(x + 1)(x + 2)^2,$$

which when expanded gives the same result.

Solution to Exercise 4.1

(a) Over $\mathbb{R}[x]$,

$$\begin{aligned}x^4 - 3x^2 + 2 &= (x - 1)(x^3 + x^2 - 2x - 2) \\ &= (x - 1)(x + 1)(x^2 - 2) \\ &= (x + 1)(x - 1)(x + \sqrt{2})(x - \sqrt{2}).\end{aligned}$$

(b) Using the answer to part (a), we see that over $\mathbb{Q}[x]$ we have

$$x^4 - 3x^2 + 2 = (x - 1)(x + 1)(x^2 - 2)$$

and $x^2 - 2$ cannot be factorised further as it has no roots in $\mathbb{Q}$.

For small finite fields such as $\mathbb{Z}_5$, it is often easiest to search exhaustively by computing $f(a)$ for every element a in the field.

(c) By inspection, over $\mathbb{Z}_5$ we see that $x = 4 \equiv -1 \pmod 5$ is a root, which gives

$$x^3 + x^2 + 2x + 2 \equiv (x + 1)(x^2 + 2) \pmod 5.$$

The quadratic $x^2 + 2$ has no roots in $\mathbb{Z}_5$, and so cannot be factorised further.

(d) Over $\mathbb{C}$, we find $f(-1) = 0$, and then using long division (or the answer to part (c)) we obtain

$$x^3 + x^2 + 2x + 2 = (x + 1)(x^2 + 2).$$

Finally, $x^2 + 2 = (x + \sqrt{2}i)(x - \sqrt{2}i)$ where $i = \sqrt{-1}$, and so the complete factorisation is

$$x^3 + x^2 + 2x + 2 = (x + 1)(x + \sqrt{2}i)(x - \sqrt{2}i).$$

(e) We find $f(1) = 0$, which gives $f(x) = (x-1)(x^3 - 5x^2 + 7x - 3)$. For the cubic, we find that $x = 1$ is again a root, which yields $f(x) = (x-1)^2(x^2 - 4x + 3)$. Finally, the quadratic has roots $x = 1$ and $x = 3$, and so

$$f(x) = (x-1)^3(x-3).$$

Solution to Exercise 4.2

(a) Let $p(x)$ be a polynomial of degree 1 in $\mathbb{Q}[x]$. Suppose that $p = ab$ for some elements $a, b \in \mathbb{Q}[x]$. Since $\deg(p) = 1$, the equation $\deg(a) + \deg(b) = \deg(p)$ implies that either $\deg(a) = 0$ or $\deg(b) = 0$. Any element in $\mathbb{Q}[x]$ of degree 0 is a unit. So p has no proper factors and is therefore irreducible.

(b) The theory of quadratic equations tells us that solutions to $f(x) = ax^2 + bx + c = 0$ are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

If $b^2 - 4ac < 0$, then $\sqrt{b^2 - 4ac} \notin \mathbb{R}$ and so $f(x) = 0$ has no solutions in $\mathbb{R}$. By the Factor Theorem, this implies that $f(x)$ has no linear factors (and therefore no proper factors) in $\mathbb{R}[x]$, and hence it is irreducible.

If $b^2 - 4ac \geq 0$, then $\sqrt{b^2 - 4ac} \in \mathbb{R}$ and so $f(x)$ has two roots in $\mathbb{R}$. This yields the proper factorisation

$$f(x) = a\left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right).$$

Solution to Exercise 4.3

(a) By the Rational Root Test, any root $\frac{p}{q}$ must satisfy $p \mid 1$, $q \mid 4$. The possible options are therefore

$$\tfrac{p}{q} = \pm 1, \pm\tfrac{1}{2}, \pm\tfrac{1}{4}.$$

Trying each in turn, we get

$$\begin{array}{ccc} f(1) = 0 & f(-1) = 6 & f(\frac{1}{2}) = 0 \\ f(-\frac{1}{2}) = \frac{3}{2} & f(\frac{1}{4}) = \frac{39}{64} & f(-\frac{1}{4}) = \frac{75}{64}. \end{array}$$

Therefore the roots of $4x^4 - 2x^3 - 2x^2 - x + 1$ in $\mathbb{Q}$ are $x = 1$ and $x = \frac{1}{2}$.

(b) The possible values that could be rational roots are $\pm 1, \pm 2, \pm 4$. Trying each in turn, we get

$$\begin{array}{ccc} f(1) = 0 & f(-1) = 6 & f(2) = 0 \\ f(-2) = 24 & f(4) = 156 & f(-4) = 300. \end{array}$$

Therefore the roots of $x^4 - x^3 - 2x^2 - 2x + 4$ in $\mathbb{Q}$ are $x = 1$ and $x = 2$.

Solution to Exercise 4.4

If $f(x)$ is reducible, then we can write $f(x) = g(x)h(x)$ in $\mathbb{Q}[x]$ for some $g, h \in \mathbb{Q}[x]$ of degree at least 1. As f is a cubic, one of g, h must have degree 1. By the Factor Theorem, it follows that $f(x)$ must have a root in $\mathbb{Q}$. Employing the Rational Root Test, the only possibilities are $\pm 1, \pm\frac{1}{2}$. Checking each of these in turn verifies that none is a root of $f(x)$, and so f has no factor of degree 1. Hence $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Solution to Exercise 4.5

First, the Rational Root Test tells us that the only possible rational numbers that could be roots of $f(x)$ are ± 1, and $f(1) = f(-1) = 1 \neq 0$. Thus f has no rational roots, and so (by the Factor Theorem) it has no linear factors.

Therefore, we investigate whether $f(x)$ can be factorised into two quadratics over $\mathbb{Z}[x]$. Suppose that

$$f(x) = (ax^2 + bx + c)(rx^2 + sx + t)$$

where $a, b, c, r, s, t \in \mathbb{Z}$. Comparing the coefficient of x^4 gives $ar = 1$, and we can assume that a is positive, so that $a = r = 1$. Next, comparing the coefficient of x^0 gives us $ct = 1$, and so $c = t = \pm 1$. Substituting the values for a, c, r and t, the coefficient of x^3 yields $b = -s$, and the coefficient of x^2 gives $-1 = \pm 1 + bs \pm 1$. Combining these to eliminate s, we find $b^2 = 1 \pm 2$.

Neither $b^2 = -1$ nor $b^2 = 3$ has a solution in the integers, and so $f(x)$ is irreducible in $\mathbb{Z}[x]$. It follows by Gauss's Lemma that $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Solution to Exercise 4.6

Remember that 0 is divisible by *every* non-zero integer.

(a) By Eisenstein's Criterion, the prime p divides every coefficient of $x^n - p$ except that of x^n. Moreover, the constant term, $-p$, is not divisible by p^2. Therefore $x^n - p$ is irreducible.

(b) First, we write $f(x) = \frac{1}{15}\left(9x^5 + 25x^4 - 15x^3 + 50x^2 + 30x - 5\right)$. We can now apply Eisenstein's Criterion with $p = 5$ to conclude that the polynomial in the brackets is irreducible. Therefore, by Lemma 4.6, so is $f(x)$.

Solution to Exercise 4.7

(a) $(x+1)^4 + 1 = x^4 + 4x^3 + 6x^2 + 4x + 2$, and so $(x+1)^4 + 1$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein's Criterion with $p = 2$. Thus by Lemma 4.17, $x^4 + 1$ is irreducible in $\mathbb{Q}[x]$.

(b) $(x+1)^8 + 1 = x^8 + 8x^7 + 28x^6 + 56x^5 + 70x^4 + 56x^3 + 28x^2 + 8x + 2$, and so $x^8 + 1$ is irreducible in $\mathbb{Q}[x]$ by Eisenstein's Criterion with $p = 2$.

In fact, it can be proved that $x^{2^n} + 1$ is irreducible in $\mathbb{Q}[x]$ for all $n \geq 0$.

Solution to Exercise 4.8

(a) $x^3 + 7x^2 - 14x + 4$ is irreducible. This can be proved using the Rational Root Test, or reducing modulo 7:

$$x^3 + 7x^2 - 14x + 4 \equiv x^3 + 4 \,(\text{mod } 7).$$

Since any proper factorisation of a cubic must include a linear factor, we now test $x = 0, 1, \ldots, 6$ in turn to see whether they are a solution of $x^3 + 4 \equiv 0 \,(\text{mod } 7)$. None is, and therefore $x^3 + 7x^2 - 14x + 4$ is irreducible in $\mathbb{Z}_7[x]$, and hence irreducible in $\mathbb{Q}[x]$ by Lemma 4.19 and Gauss's Lemma.

(b) This polynomial is not irreducible in $\mathbb{Q}[x]$. The Rational Root Test can be used to prove that $x^4 + 4$ has no roots, and therefore no linear factors. It follows that any proper factor must be quadratic. Without loss of generality, we can assume $x^4 + 4$ to be the product of two monic quadratics, that is, $x^4 + 4 = (x^2 + ax + b)(x^2 + cx + d)$. Comparing coefficients then leads to the factorisation

$$x^4 + 4 = (x^2 + 2x + 2)(x^2 - 2x + 2).$$

As there are no linear factors, this factorisation must be complete.

Solution to Exercise 4.9

(a) The irreducible polynomials are

$$\begin{array}{cccc} x, & x+1, & x^2+x+1, & x^3+x+1, \\ x^3+x^2+1, & x^4+x+1, & x^4+x^3+1, & x^4+x^3+x^2+x+1. \end{array}$$

The quadratic and cubic polynomials that are reducible (and therefore not listed above) must have a linear factor, and so either $x = 0$ or $x = 1$ is a root.

Notice that any quartic that is reducible but has no linear factors must be a product of irreducible quadratic polynomials. In $\mathbb{Z}_2[x]$, there is only one irreducible quadratic, namely $x^2 + x + 1$. Thus $(x^2 + x + 1)^2 \equiv x^4 + x^2 + 1 \,(\text{mod } 2)$ is reducible, but all other quartics are irreducible or have a linear factor modulo 2.

(b) (i) $x^4 - 3x^3 + 12x^2 - 2x + 5 \equiv x^4 + x^3 + 1 \,(\text{mod } 2)$ and so is irreducible in $\mathbb{Z}_2[x]$, and hence in $\mathbb{Q}[x]$.

(ii) $3x^4 - x^3 + 71x^2 - 9x + 1 \equiv x^4 + x^3 + x^2 + x + 1 \,(\text{mod } 2)$ and so is irreducible in $\mathbb{Z}_2[x]$, and thus in $\mathbb{Q}[x]$.

Chapter 12

Fermat's Last Theorem and unique factorisation

Introduction

In the last chapter, we began our exploration of ring theory by looking at the specific case of polynomial rings over fields. We learnt how to carry out polynomial division over arbitrary fields and compute highest common factors, and we then established conditions for irreducibility of polynomials.

In this chapter, we will extend our use of the word 'irreducible' to arbitrary rings and use this to explore how we may generalise the *Fundamental Theorem of Arithmetic*; that is, we will ask when we can factorise an element of a ring R into irreducible elements, and when this can be done uniquely. As you will see, this story of unique factorisation relates strongly to the solutions of certain Diophantine equations and it is with the story of Fermat's Last Theorem that our investigation begins.

Do not confuse this with the Fundamental Theorem of Algebra, which concerns irreducible polynomials over $\mathbb{C}[x]$.

Chapter guide

This chapter has four sections. The first section, although setting the theme for the rest of the chapter, does not contain mathematics used by the later sections. The three later sections, however, do build on each other, and they contain material that will be used in Book E. These three sections are of roughly equal length, although the second section contains fewer technical arguments than the final two.

1 Fermat's Last Theorem and Diophantine equations

Fermat's refusal to publish his discoveries led to one of the great stories in the history of mathematics. As well as writing letters to his friends about various results, he formed the habit of jotting notes in the margins of his reference books. Some five years after his death, his copy of Bachet's *Diophantus* was found to contain mention of many results from number theory. On the page dealing with the Pythagorean equation, Fermat had added the following in the margin.

> On the other hand it is impossible to write a cube as the sum of two cubes, a fourth power as the sum of two fourth powers and, in general, any power higher than the second as the sum of two similar powers. I have discovered a truly marvellous proof of this, but the margin is too small to contain it.

Here, Fermat was claiming to have proved that the Diophantine equation $x^n + y^n = z^n$ has no positive integer solutions for $n \geq 3$. No evidence of

how Fermat reached his conclusions survives, and it took another three and a half centuries of continual efforts to solve the problem. As this was the last of Fermat's results awaiting proof or refutation, it became known colloquially as *Fermat's Last Theorem*. Strictly speaking, until its proof it should not have been called a theorem, and you will find it referred to as Fermat's Conjecture in some older texts.

Consider the equation $x^n + y^n = z^n$. When $n = 2$, this is the Pythagorean equation, and we will classify the integer solutions to this in the next subsection. For $n \geq 3$, if the exponent n is composite, say $n = kr$, then the equation can be rewritten as

$$(x^k)^r + (y^k)^r = (z^k)^r.$$

It follows that if $x = x_0$, $y = y_0$, $z = z_0$ is a solution of $x^n + y^n = z^n$ then

$$x = x_0^k, \quad y = y_0^k, \quad z = z_0^k$$

is a solution of $x^r + y^r = z^r$. Conversely, if it can be established that $x^r + y^r = z^r$ has no positive solutions then it follows that $x^n + y^n = z^n$ must have no positive solutions.

Now, every integer $n \geq 3$ is divisible by some odd prime or by 4. The proof of Fermat's Last Theorem can therefore be reduced to proving two cases:

1. $x^4 + y^4 = z^4$ has no positive solution
2. $x^p + y^p = z^p$ has no positive solution for any odd prime p.

In this section, after we have solved the Pythagorean equation, we will present the method of 'infinite descent' and use this, as Fermat himself did, to prove the case $n = 4$. The real challenge, however, lay in proving Fermat's Last Theorem for odd prime exponents, and we will pick up this story at the start of the next section.

1.1 Pythagorean triples

Pythagoras' Theorem, which states that the sum of the squares of the two shorter sides of a right-angled triangle is equal to the square of the hypotenuse (Figure 1.1), is arguably the most celebrated theorem in mathematics. But in addition to being a classical theorem of geometry, it presents challenges for the number theorist when the underlying equation $x^2 + y^2 = z^2$ is considered as a Diophantine equation. We refer to this as the **Pythagorean equation**.

We will continue the geometric analogy by referring to the value of z in any solution as being the **hypotenuse**.

Of course, unlike the equation in Fermat's Last Theorem, the Pythagorean equation $x^2 + y^2 = z^2$ has non-trivial solutions in the natural numbers, including the well-known one that has the smallest value of z, namely $3^2 + 4^2 = 5^2$. It is also immediate from this one solution that the equation has infinitely many solutions since, for any positive integer k,

$$(3k)^2 + (4k)^2 = (5k)^2.$$

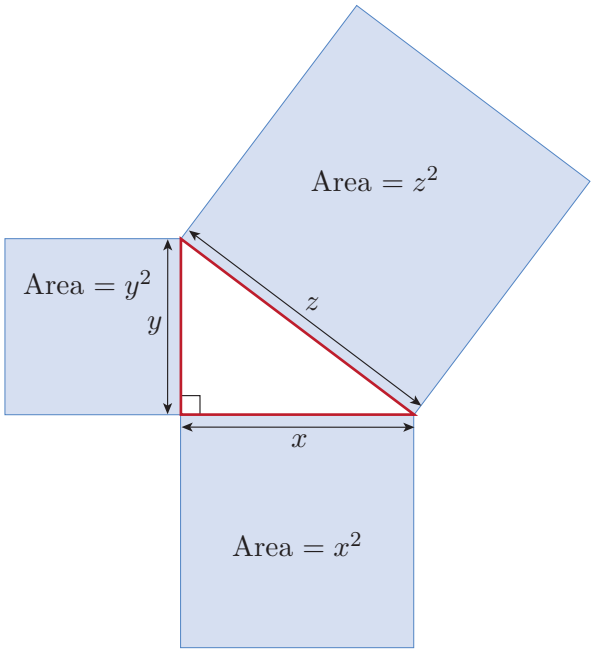


Figure 1.1 Pythagoras' Theorem: $x^2 + y^2 = z^2$

But there is a sense in which this infinite family of solutions is really just one solution. Geometrically, the corresponding triangles are similar; we have the one basic triangle with sides 3, 4 and 5, and the others are obtained by scaling each side by factor k.

Before progressing further, let us introduce some terminology.

Definition 1.1 *Pythagorean triples*

A **Pythagorean triple** is a triple (x, y, z) of positive integers such that $x^2 + y^2 = z^2$.

The triple is said to be **primitive** if $\text{hcf}(x, y) = 1$.

A triple is ordered so that the hypotenuse is always the third member.

Notice that the condition $\text{hcf}(x, y) = 1$ guarantees that x and y have no common factor greater than 1, and consequently the primitive triple (x, y, z) cannot be a scaling of some smaller triple. Moreover, the condition $\text{hcf}(x, y) = 1$ carries the implications that $\text{hcf}(x, z) = \text{hcf}(y, z) = 1$, since from the equation $x^2 + y^2 = z^2$ it is readily shown that any common divisor of two of the variables must divide the third.

Exercise 1.1

Which, if any, of the following are Pythagorean triples? For each Pythagorean triple, decide whether or not it is primitive.

(a) $(10, 8, 6)$

(b) $(12, 9, 15)$

(c) $(6, 7, 8)$

(d) $(5, 12, 13)$

(e) $(24, 33, 41)$

We now have two primitive Pythagorean triples, namely $(3, 4, 5)$ and $(5, 12, 13)$. Are there any more? The Babylonians, some 3500 years ago, were aware of many primitive solutions, but Heath, in Book I of his translation of Euclid's *Elements*, credits Pythagoras with the first infinite family of solutions given by

$$x = 2k + 1, \quad y = 2k^2 + 2k, \quad z = 2k^2 + 2k + 1, \quad \text{for any integer } k \geq 1.$$

The first complete solution of the Pythagorean equation appeared in Euclid's *Elements*. We aim to reproduce that solution here, but first there are a few points to be clarified.

If (x, y, z) is any Pythagorean triple then x and y cannot both be odd. For if they were then

$$z^2 = x^2 + y^2 \equiv 1 + 1 \equiv 2 \,(\text{mod } 4),$$

which is impossible as all squares are congruent modulo 4 to either 0 or 1. It follows that in a primitive triple x and y have opposite parity, because the further condition that $\text{hcf}(x, y) = 1$ ensures that they cannot both be even.

The **parity** of a pair of integers is defined as follows. If both are odd or both are even, they are said to have the *same* parity; otherwise they are said to have *opposite* parity.

If (x, y, z) is a primitive Pythagorean triple then so too is (y, x, z), because the x and y values can certainly be interchanged. But, to all intents and purposes, the triples (x, y, z) and (y, x, z) lead to what is really the same solution of the Pythagorean equation. To overcome the need to distinguish between these two equivalent solutions, we choose to write the even member first in any primitive triple. We adopt the following convention.

Convention for Pythagorean triples

If (x, y, z) is any primitive Pythagorean triple, then x is even while y and z are both odd.

With the preparation now complete, we are ready for the main theorem of this subsection.

Theorem 1.2 *Primitive solutions of the Pythagorean equation*

The primitive Pythagorean triples are the triples

$$(2mn, m^2 - n^2, m^2 + n^2),$$

where m and n are coprime natural numbers of opposite parity with $m > n$.

Before embarking on the proof, let us clarify the claim of the theorem by looking at an example of its use.

Worked Exercise 1.3

Find all primitive Pythagorean triples $(60, y, z)$.

Solution

We seek coprime integers m and n such that $2mn = 60$, m and n have opposite parity and $m > n$. There are four pairs of positive integers satisfying $mn = 30$ with $m > n$, and in all four cases m and n are coprime and have opposite parity:

$m = 30,\ n = 1,$ giving $(60, 899, 901)$;
$m = 15,\ n = 2,$ giving $(60, 221, 229)$;
$m = 10,\ n = 3,$ giving $(60, 91, 109)$;
$m = \ 6,\ n = 5,$ giving $(60, 11, 61)$.

These four are the only primitive Pythagorean triples $(60, y, z)$.

Proof of Theorem 1.2 We must first show that the given triple is a primitive Pythagorean triple. Substituting $x = 2mn$, $y = m^2 - n^2$ and $z = m^2 + n^2$ into the Pythagorean equation, we have

$$\begin{aligned} x^2 + y^2 &= (2mn)^2 + (m^2 - n^2)^2 \\ &= m^4 + 2m^2n^2 + n^4 = (m^2 + n^2)^2 = z^2. \end{aligned}$$

So this is indeed a Pythagorean triple. To establish the primitive property, suppose to the contrary that p is a prime factor of both $m^2 - n^2$ and $m^2 + n^2$. Then p divides both the sum $(m^2 + n^2) + (m^2 - n^2) = 2m^2$ and the difference $(m^2 + n^2) - (m^2 - n^2) = 2n^2$. But $p \neq 2$ (since y is odd) and so we conclude that p divides m and p divides n. As m and n are given to be coprime we have the required contradiction, and so the triple is primitive.

Conversely, we have to show that every primitive Pythagorean triple is of the stated form. With this goal in mind, suppose that (x, y, z) is *any* primitive Pythagorean triple. As y and z are both odd, we can define integers

$$s = \frac{z+y}{2} \text{ and } t = \frac{z-y}{2}.$$

We note that s and t are coprime because any common factor would also divide $s + t = z$ and $s - t = y$.

Then from the Pythagorean equation

$$x^2 = z^2 - y^2 = (z+y)(z-y) = 2s \times 2t$$

we obtain

$$\left(\frac{x}{2}\right)^2 = st.$$

As x is known to be even, this equation confirms that the product st is a perfect square. Therefore, since s and t are coprime, they must each be squares. So if we now write $s = m^2$ and $t = n^2$ and substitute back, we get

This follows from the work of Book A, Chapter 2, but see also Exercise 1.2(c) at the end of this subsection.

$$x^2 = 4st = 4m^2n^2, \quad \text{giving } x = 2mn;$$
$$y = s - t = m^2 - n^2;$$
$$z = s + t = m^2 + n^2.$$

Finally, note that $\text{hcf}(m, n) = 1$, because any common factor of m and n would necessarily be a common factor of the coprime integers $s = m^2$ and $t = n^2$. Moreover, m and n must have opposite parity for otherwise y and z would be even.

The triple (x, y, z) has thus been expressed in the required form. ■

All Pythagorean triples can be obtained by scaling primitive ones; that is, any Pythagorean triple is of the form (kx, ky, kz), where (x, y, z) is a primitive triple. Notice that our convention for primitive triples carries over to all triples, and so the first member of the triple is always the multiple of the even member in the underlying primitive triple. With this clarified, we can now record the following corollary.

Corollary 1.4 *All solutions of the Pythagorean equation*

The Pythagorean triples (x, y, z) are given by

$$x = 2kmn, \quad y = k(m^2 - n^2), \quad z = k(m^2 + n^2),$$

where $k \geq 1$ is any integer and m and n are coprime natural numbers with opposite parity and $m > n$.

The fact that m, n are coprime implies that there are no Pythagorean triples (x, y, z) with $x = y$.

In the next worked exercise we make use of Theorem 1.2 to begin an enumeration of all the primitive Pythagorean triples.

Worked Exercise 1.5

There are 16 primitive Pythagorean triples (x, y, z) with hypotenuse $z < 100$. Find them.

Solution

We find primitive triples by listing pairs of coprime integers m and n that have opposite parity and with $m > n$. For each such pair, the corresponding triple is calculated from the formulas in Theorem 1.2. For example, taking $m = 7$, each of the values $n = 2$, 4 and 6 will lead to a primitive triple.

For $m = 7$ and $n = 1$, 3 or 5 the resulting triple will be Pythagorean but will not be primitive, because m and n have the same parity.

The condition that $z < 100$ amounts to $m^2 + n^2 < 100$. Hence $m \leq 9$ and, for example, the triple resulting from $m = 8$, $n = 7$ is not wanted because its hypotenuse exceeds 100. The 16 solutions are as shown in Table 1.1.

Table 1.1 Primitive triples with sides not exceeding 100

m	n	triple	m	n	triple
2	1	$(4, 3, 5)$	7	2	$(28, 45, 53)$
3	2	$(12, 5, 13)$	7	4	$(56, 33, 65)$
4	1	$(8, 15, 17)$	7	6	$(84, 13, 85)$
4	3	$(24, 7, 25)$	8	1	$(16, 63, 65)$
5	2	$(20, 21, 29)$	8	3	$(48, 55, 73)$
5	4	$(40, 9, 41)$	8	5	$(80, 39, 89)$
6	1	$(12, 35, 37)$	9	2	$(36, 77, 85)$
6	5	$(60, 11, 61)$	9	4	$(72, 65, 97)$

Exercise 1.2

(a) Find all primitive Pythagorean triples $(72, y, z)$.

(b) Find all Pythagorean triples, primitive or not, in which one of the sides is 30.

(c) Give a proof of the step assumed in Theorem 1.2 that if a product of two coprime integers is a square, then each of the integers must be square.

1.2 The equation $x^4 + y^4 = z^4$

In the remainder of this section we will confine our attention to the special case of Fermat's Last Theorem for exponent 4 and some related Diophantine equations. We will use Fermat's method and, just as he did, prove the stronger result that $x^4 + y^4 = z^2$ has no positive solution. The proof uses our solution to the Pythagorean equation together with a powerful technique known as Fermat's **method of infinite descent**.

The idea behind the method of infinite descent as applied to this problem is as follows. We establish that from any positive solution $x = x_k$, $y = y_k$, $z = z_k$ of the Diophantine equation $x^4 + y^4 = z^2$ we can always find another 'smaller' positive solution $x = x_{k+1}$, $y = y_{k+1}$, $z = z_{k+1}$, where by smaller we mean that $0 < z_{k+1} < z_k$. This process of being able to go from one positive solution to a smaller one is the key stage of the method, and is called the **descent step**.

Now suppose that the equation does have a positive solution $x = x_1$, $y = y_1$, $z = z_1$. The established descent step would then tell us that, from this, we have a smaller positive solution, and from this second solution a still smaller third positive solution, from which we get a smaller fourth positive solution, and so on. Focusing on the z-values, we have an unending decreasing sequence of positive integers

$$z_1 > z_2 > z_3 > z_4 > \cdots > 0.$$

As there are only finitely many positive integers that are less than z_1, this is plainly impossible. Hence the one assumption that we have made, namely that one positive solution exists, is contradicted.

Infinite descent bears some resemblance to mathematical induction. The descent step, like the induction step, sets up an unending chain of implications of the form 'if one is true then so too is the next'. The difference is that this time we know that no such infinite chain of true statements can exist and we are forced to conclude that the chain cannot be initiated; it has no basis.

We have chosen to focus on the variable z here. But notice that descent could be achieved on any of the variables. For example, if we established that any positive solution of the equation must give rise to another positive solution with smaller value for y, then the conclusion that no solution can exist follows in the same way.

What is really being brought into play here is the Well-Ordering Principle from Book A, Chapter 1, which asserts that any non-empty set of positive integers must have a least member. The descent step appears to construct a set of positive integers that does not have a least member, and so it has to be the empty set.

Theorem 1.6

The Diophantine equation $x^4 + y^4 = z^2$ has no positive solutions.

The proof of Theorem 1.6 and that of Theorem 1.9, which follows shortly, involve some tricky algebra. You are not necessarily expected to master the details of these proofs, but you should read each of them carefully to see how the method of infinite descent is used.

Proof Suppose to the contrary that $x^4 + y^4 = z^2$ has a positive solution $x = x_1$, $y = y_1$, $z = z_1$. If $\mathrm{hcf}(x_1, y_1) = d > 1$ then putting $x_1 = dx'$ and $y_1 = dy'$ gives $z_1^2 = d^4(x'^4 + y'^4)$, from which it follows that d^2 divides z_1 so that $z_1 = d^2 z'$, for some integer $z' < z_1$. But then $x'^4 + y'^4 = z'^2$, where $\mathrm{hcf}(x', y') = 1$ and $z' < z_1$. This argument shows that we may assume that $\mathrm{hcf}(x_1, y_1) = 1$, for otherwise any common divisor can first be cancelled leaving another, smaller solution to the same equation.

Rather than use the general suffix k as in the preamble, we establish the descent step here by showing that any positive solution x_1, y_1, z_1 must lead to a positive solution x_2, y_2, z_2 in which $0 < z_2 < z_1$.

Writing the equation as $(x_1^2)^2 + (y_1^2)^2 = z_1^2$, we observe that (x_1^2, y_1^2, z_1) is a primitive Pythagorean triple and so, by Theorem 1.2,

If needed we may interchange x_1 and y_1 to ensure that x_1 is even.

$$x_1^2 = 2mn, \quad y_1^2 = m^2 - n^2, \quad z_1 = m^2 + n^2,$$

where m and n are coprime positive integers of opposite parity with $m > n$. In fact n must be even and m odd, for otherwise we have

$$y_1^2 = m^2 - n^2 \equiv 0 - 1 \equiv 3 \,(\mathrm{mod}\ 4),$$

which is impossible because any square is congruent to either 0 or 1 modulo 4.

Putting $n = 2r$, the equation for x_1 becomes

$$\left(\frac{x_1}{2}\right)^2 = mr,$$

where m and $r = \frac{n}{2}$ are coprime. Hence m and r are each squares, say $m = s^2$ and $r = t^2$.

Returning to the equation $y_1^2 = m^2 - n^2$, we see that (n, y_1, m) is a primitive Pythagorean triple and so

$$n = 2uv, \quad y_1 = u^2 - v^2, \quad m = u^2 + v^2,$$

where u and v are coprime positive integers of opposite parity with $u > v$.

Now $n = 2r = 2t^2$, and so the first of these equations becomes $uv = t^2$. It follows that u and v are each squares, say $u = x_2^2$ and $v = y_2^2$. Feeding these facts into the equation $m = u^2 + v^2$ gives

$$x_2^4 + y_2^4 = m = s^2 (= z_2^2, \text{ say}),$$

revealing another solution of the original equation. But this solution is 'smaller' since

$$0 < z_2 = s \leq m < m^2 + n^2 = z_1.$$

This completes the descent step; any positive solution gives rise to a smaller positive one. As this is impossible, the assumption that there exists a positive solution is contradicted. ■

As any fourth power is necessarily a square, any positive solution of $x^4 + y^4 = z^4$ would contradict the result of Theorem 1.6. We deduce the following.

Corollary 1.7

The Diophantine equation $x^4 + y^4 = z^4$ has no positive solutions.

The method of infinite descent used in the proof of Theorem 1.6 can be applied to many Diophantine equations, almost invariably with the purpose of showing that the equation has no positive solutions. Here is a simpler example.

Worked Exercise 1.8

Show that the Diophantine equation $x^3 + 3y^3 = 9z^3$ has no positive solutions.

Solution

Aiming for a contradiction, suppose that $x_1^3 + 3y_1^3 = 9z_1^3$, where x_1, y_1 and z_1 are positive integers. As the prime 3 divides the right-hand side and one of the terms on the left of this equation, it must divide the remaining term. That is, 3 divides x_1^3 and so 3 divides x_1. This means we can find a positive integer x_2 so that $x_1 = 3x_2$, and this gives us

$$27x_2^3 + 3y_1^3 = 9z_1^3.$$

We can divide this equation by 3, to obtain

$$9x_2^3 + y_1^3 = 3z_1^3.$$

Repetition of the same reasoning now shows that 3 divides y_1, and so there is a positive integer y_2 with $y_1 = 3y_2$. This gives

$$9x_2^3 + 27y_2^3 = 3z_1^3; \quad \text{that is, } 3x_2^3 + 9y_2^3 = z_1^3.$$

This time we have 3 dividing z_1, and putting $z_1 = 3z_2$ for some positive integer z_2, the equation now becomes

$$3x_2^3 + 9y_2^3 = 27z_2^3; \quad \text{that is, } x_2^3 + 3y_2^3 = 9z_2^3.$$

At this point we have reached a second positive solution, $x = x_2$, $y = y_2$, $z = z_2$ of the original equation with $z_2 < z_1$. The descent step is therefore complete and the required contradiction is established. Thus $x^3 + 3y^3 = 9z^3$ has no positive solutions.

Exercise 1.3

Show that the Diophantine equation $x^4 + 2y^4 = 4z^4$ has no positive solutions.

The following exercise requires a slight variant on Fermat's method of infinite descent.

Exercise 1.4

(a) Show that if $x^2 + y^2 + z^2 \equiv 0 \pmod 4$ then each of x, y and z is even.

(b) Show that in any positive solution $x = x_1$, $y = y_1$, $z = z_1$ of the Diophantine equation $x^2 + y^2 + z^2 = 2xyz$, each of x_1, y_1 and z_1 must be even. Putting $x_1 = 2x_2$, $y_1 = 2y_2$ and $z_1 = 2z_2$, show that x_2, y_2 and z_2 are also even positive integers.

(c) Prove that the Diophantine equation $x^2 + y^2 + z^2 = 2xyz$ has no positive solution by constructing from such a solution a strictly decreasing infinite sequence of even positive integers.

1.3 Related Diophantine equations

Fermat went on to apply his method of infinite descent to prove that the related equation $x^4 - y^4 = z^2$ has no positive solution. We will, however, prove this result as a corollary of the following theorem.

Theorem 1.9

The Diophantine equation $x^4 + 4y^4 = z^2$ has no positive solution.

Proof Suppose that x_1, y_1 and z_1 are positive integers with $x_1^4 + 4y_1^4 = z_1^2$, that is, $(x, y, z) = (x_1, y_1, z_1)$ is a particular solution. Exactly as in the proof of Theorem 1.6 we may assume that $\text{hcf}(x_1, y_1) = 1$, for otherwise cancellation of any common divisor leads to a smaller solution of the same equation.

With infinite descent in mind, we wish to show that the existence of this one positive solution necessarily gives rise to one with a smaller z-value. We observe that x_1 and z_1 are of the same parity, and we first consider the case in which they are both even. Substituting $x_1 = 2x'$ and $z_1 = 2z'$ gives

$$(2x')^4 + 4y_1^4 = (2z')^2; \quad \text{that is, } y_1^4 + 4x'^4 = z'^2,$$

which is another positive solution of the equation with $z' < z_1$.

The solution is $(x, y, z) = (y_1, x', z')$.

It remains to show that a solution in which x_1 and z_1 are both odd must likewise lead to another with smaller z-value. Writing the equation as $(x_1^2)^2 + (2y_1^2)^2 = z_1^2$, we recognise that $(2y_1^2, x_1^2, z_1)$ is a primitive

Pythagorean triple and so, according to Theorem 1.2, there are coprime positive integers m and n of opposite parity and with $m > n$ such that

$$2y_1^2 = 2mn, \quad x_1^2 = m^2 - n^2, \quad z_1 = m^2 + n^2.$$

From $x_1^2 = m^2 - n^2$ we see that m is odd and n is even, for the other way round would give $m^2 - n^2 \equiv 3 \pmod 4$ and this cannot be a square. And from $y_1^2 = mn$, the coprime integers m and n are each squares and so we can write $m = a^2$ and $n = (2b)^2$.

The triple is primitive because $\text{hcf}(m, n) = 1$.

Looking again at $x_1^2 = m^2 - n^2$ with $\text{hcf}(m, n) = 1$, we now see that (n, x_1, m) is another primitive Pythagorean triple, and so there exist coprime positive integers s and t, of opposite parity and with $s > t$ such that

$$n = 2st, \quad x_1 = s^2 - t^2, \quad m = s^2 + t^2.$$

Substituting $n = (2b)^2$ into the first of these equations gives $2b^2 = st$. Now one of s or t is even. If it is s, say $s = 2r$, then $b^2 = rt$ and as r and t are coprime each is a square; say $r = u^2$ (so that $s = 2u^2$) and $t = v^2$. The alternative that t is even leads, in exactly the same way, to $s = v^2$ and $t = 2u^2$. In either case, substituting into $m = s^2 + t^2$ gives

$$v^4 + 4u^4 = s^2 + t^2 = m = a^2.$$

Thus we have another solution of the original equation. Moreover, putting $z_2 = a$, we have

$$0 < z_2 = a \leq m < m^2 + n^2 = z_1,$$

and so the z-value in this new solution is a smaller positive integer.

This completes the descent step and the proof. ■

Corollary 1.10

The Diophantine equation $x^4 - y^4 = z^2$ has no positive solution.

Proof Suppose that x_1, y_1 and z_1 are positive integers with $x_1^4 - y_1^4 = z_1^2$. Squaring both sides of this equation and rearranging gives

$$z_1^4 + 4(x_1 y_1)^4 = (x_1^4 + y_1^4)^2.$$

Thus $x = z_1$, $y = x_1 y_1$, $z = x_1^4 + y_1^4$ is a positive solution of $x^4 + 4y^4 = z^2$. But this contradicts Theorem 1.9 and so the equation $x^4 - y^4 = z^2$ has no positive solution. ■

We finish with a problem for you to attempt. The solution to this problem first appeared in the margin of Fermat's copy of *Diophantus*. We refer to a right-angled triangle associated with a Pythagorean triple as a **Pythagorean triangle**.

Exercise 1.5

Let (x, y, z) be a Pythagorean triple. The area of the associated Pythagorean triangle is $\frac{1}{2}xy$, and for this to be a square we need $xy = 2n^2$ for some integer n.

Show that if there is a positive simultaneous solution to this equation and to $x^2 + y^2 = z^2$, then there is a positive solution of $a^4 - b^4 = c^2$, and hence deduce that no Pythagorean triangle can have an area that is a square.

2 Integral domains

Continuing the discussion at the start of Section 1, we now have proof that the equation $x^4 + y^4 = z^4$ has no integer solutions, and therefore the proof of Fermat's Last Theorem would follow if it could be shown that $x^p + y^p = z^p$ has no solutions where p is an odd prime.

Individual cases were resolved over a number of years: $p = 3$ was proved by Euler in 1770, half a century later $p = 5$ was proved independently by Dirichlet (1805–1859) and Legendre, and $p = 7$ was proved by Lamé (1795–1870) in 1835. Meanwhile, Germain (1776–1831) proved a property about the divisibility of the product xyz of a solution to $x^p + y^p = z^p$, and this enabled her to show that there can be no solution for all odd primes $p \leq 197$.

Lamé thought he had made the breakthrough in 1847. In attempting to solve $x^p + y^p = z^p$, he used the ring

$$\mathbb{Z}[\omega] = \{x_0 + x_1\omega + x_2\omega^2 + \cdots + x_{p-1}\omega^{p-1} : x_i \in \mathbb{Z}\}$$

where ω is a complex primitive pth root of 1, that is, $\omega^p = 1$. Lamé showed that the left-hand side of $x^p + y^p = z^p$ can be factorised as follows:

$$x^p + y^p = (x + y)(x + \omega y)(x + \omega^2 y) \cdots (x + \omega^{p-1} y).$$

Applying Fermat's method of infinite descent, he showed that there can be no trivial solution in this number system, and therefore no positive solution. Unfortunately, this relied on the above factorisation of $x^p + y^p$ being unique, which is not in general true: the equivalent of the Fundamental Theorem of Arithmetic does not always hold in $\mathbb{Z}[\omega]$. For the rest of the chapter, we will be exploring this idea of unique factorisation in general rings, and how this can be used to solve some other Diophantine equations.

In fact, the smallest p for which there is no analogue to the Fundamental Theorem of Arithmetic in $\mathbb{Z}[\omega]$ is $p = 23$.

2.1 Units and associates

Our exploration into unique factorisation begins by picking up the story of rings, zero divisors and units from Chapter 11. For polynomial rings over fields, we have already defined a number of concepts that we will need here, and so some of the terms that we will define should already be familiar. Here, however, we are going to be considering arbitrary commutative rings, and not just polynomial rings. Note that, as in Chapter 11, all the rings we consider will be commutative rings. However, we will occasionally use the term 'commutative' to emphasise the fact that a definition or result does not apply more generally.

We begin by recalling that an element a of a ring R is a **unit** if there exists $b \in R$ for which $a \cdot b = 1$. We saw in Subsection 1.2 of Chapter 11 that the units of $\mathbb{Z}$ are 1 and -1. In other rings, the computation of units can be much harder, as the following example demonstrates.

Example 2.1 *Units*

Consider $\mathbb{Z}[\sqrt{2}]$, the ring of real numbers of the form $a + b\sqrt{2}$ with $a, b \in \mathbb{Z}$. Set $u = 1 + \sqrt{2}$. Then u is a unit, because

$$(1 + \sqrt{2})(-1 + \sqrt{2}) = -1 + \sqrt{2} - \sqrt{2} + \sqrt{2}\sqrt{2} = -1 + 2 = 1.$$

Now, u is a real number that is greater than 1, and so all of the numbers $u, u^2, u^3, \ldots$ are distinct. Moreover, u^k is a unit for every $k \geq 1$, because

$$\begin{aligned}(1 + \sqrt{2})^k(-1 + \sqrt{2})^k &= (1 + \sqrt{2})^{k-1}\left((1 + \sqrt{2})(-1 + \sqrt{2})\right)(-1 + \sqrt{2})^{k-1} \\ &= (1 + \sqrt{2})^{k-1}(-1 + \sqrt{2})^{k-1} \\ &= \cdots = 1.\end{aligned}$$

This shows us, therefore, that $\mathbb{Z}[\sqrt{2}]$ has infinitely many units.

Since $u > 1$, it follows that $1 < u < u^2 < u^3 < \cdots$, which implies that each u^k is distinct. In fact, the only units of $\mathbb{Z}[\sqrt{2}]$ are of the form $1, -1, (1 + \sqrt{2})^k$ or $(-1 + \sqrt{2})^k$, for $k \geq 1$, but we will not prove this here.

How did we end up with infinitely many units in the above example? We managed to find a unit $u = 1 + \sqrt{2}$ for which every power of u is also a unit. This, of course, holds more generally.

Lemma 2.2

Let R be a commutative ring, and u a unit of R. Then u^k is a unit of R for all $k \geq 1$.

Proof Since u is a unit, there exists u^{-1} such that $uu^{-1} = 1$ in R. The inverse of u^k is then $(u^{-1})^k$, since

$$u^k \cdot (u^{-1})^k = u^{k-1}(uu^{-1})(u^{-1})^{k-1} = \cdots = uu^{-1} = 1,$$

and hence u^k is a unit. ■

In $\mathbb{Z}$, we know that the units are ± 1. Now, $(-1)^2 = 1$ (and, of course, $1^2 = 1$), which explains why the above lemma does not mean that $(-1)^k$ creates an infinite list of units in $\mathbb{Z}$.

Knowing about the units of a given ring is important if we are to extend our concepts of factorisation. For example, in $\mathbb{Z}$ we do not want to regard the two factorisations $6 = 2 \cdot 3 = (-2) \cdot (-3)$ as being different: the number -2 is really just the number 2 multiplied by the unit -1, and units are always going to let us do this. This motivates us to extend a concept that we used for polynomial rings (Definition 2.11 in Chapter 11) to a more general setting.

Definition 2.3 *Associate*

Let R be a commutative ring. An **associate** of $a \in R$ is any element that can be written as ua where u is a unit in R.

Exercise 2.1

Let $d > 1$ be a non-square integer.

(a) Show that if $a, b \in \mathbb{Z}$ is a solution to the equation $a^2 - db^2 = 1$, then $a + b\sqrt{d}$ is a unit in $\mathbb{Z}[\sqrt{d}]$.

(b) Hence show that whenever $a^2 - db^2 = 1$ has a solution with $a, b > 0$, then $\mathbb{Z}[\sqrt{d}]$ has infinitely many units.

The equation $a^2 - db^2 = 1$ became known as **Pell's equation** when Euler confused John Pell (1611–1685) with another English mathematician, William Brouncker (1620–1684), who had been studying the equation.

In fact, Pell's equation $a^2 - db^2 = 1$ has infinitely many non-trivial solutions (a, b) for every square-free integer $d > 1$. Although we will not prove this fact here, combining it with the exercise above implies that $\mathbb{Z}[\sqrt{d}]$ has infinitely many units whenever $d > 1$. By contrast, $\mathbb{Z}[\sqrt{-1}]$ has four units, namely $1, -1, i$ and $-i$, and the only units in $\mathbb{Z}[\sqrt{d}]$ for $d < -1$ are -1 and 1.

2.2 Factors, irreducibles, primes and integral domains

When trying to define 'factorisation' for an element $a \in R$, we want to exclude factors that are units, since we can always write $a = uu^{-1}a$ for a unit u and its inverse. Following a similar concept to the one used in our work with polynomial rings, in the next definition we therefore exclude units from being considered proper factors.

We will occasionally use the shorthand notation $a \mid c$ to mean 'a divides c'.

Definition 2.4 *Factor*

Let R be a commutative ring. We say that $a \in R$ is a **factor** or **divisor** of $c \in R$ (or a **divides** c) if there is $b \in R$, such that $c = ab$. If c is non-zero and neither a nor b are units then they are **proper factors** of c.

Note that this definition of factor and proper factor agrees with what we are used to in $\mathbb{Z}$: a proper factor of an integer $n \in \mathbb{Z}$ is any factor of n other than n itself, its associate $-n$ and the two units 1 and -1. Similarly, in the polynomial ring $F[x]$, a proper factor of f must have degree at least one and at most $\deg(f) - 1$, so that it is neither a unit nor an associate.

However, you may have noticed that, unlike in the cases of $\mathbb{Z}$ and $F[x]$, we have defined what it means for one element to divide another without first describing a division algorithm. We will explore why we did this in the next section.

With 'factor' defined, our next step is to generalise the term 'irreducible', which we encountered in Chapter 11. We will present the definition here slightly differently, but it is easy to see that this is equivalent to the approach used in the last chapter.

Definition 2.5 *Irreducible*

Let R be a commutative ring. We say that $a \in R$ is **irreducible** if it is not a unit, it is not zero and whenever $a = bc$ for $b, c \in R$, then one of b or c is a unit.

In the integers, we used the term 'prime' instead of irreducible, and you may be wondering why the terminology has changed. The reason is that in the general context of studying rings, the term 'prime' is used to define something related, but subtly different.

Definition 2.6 *Prime*

Let R be a commutative ring. A non-zero, non-unit element p of R is **prime** if whenever $p \mid ab$ for $a, b \in R$, then either $p \mid a$ or $p \mid b$.

We will soon see that Definitions 2.5 and 2.6 coincide in $\mathbb{Z}$ (and define the 'prime numbers' that we are familiar with). In the more general setting of rings, understanding the distinction between irreducible and prime is key to our exploration of unique factorisation. When do they define different objects? There are really two questions here: when are irreducible elements prime, and when are prime elements irreducible? We will consider the second of these questions in the next subsection. To answer the first, we consider an example.

Example 2.7 *A prime element that is not irreducible*

Take the ring $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$, of integers with addition and multiplication modulo 6. Then the element 2 is prime in $\mathbb{Z}_6$. To see this, we write out the multiplication table for $\mathbb{Z}_6 \setminus \{0\}$.

	1	2	3	4	5
1	1	2	3	4	5
2	2	4	0	2	4
3	3	0	3	0	3
4	4	2	0	4	2
5	5	4	3	2	1

Reading along the row for 2, we see that 2 divides the numbers 2, 4 and 0. Now, we can write 4 as the product of two elements in four ways:

$$4 = 1 \cdot 4 = 2 \cdot 2 = 2 \cdot 5 = 4 \cdot 4.$$

In each of these four products, we find that 2 divides one of the two factors. We can do the same with 2:

$$2 = 1 \cdot 2 = 2 \cdot 4 = 4 \cdot 5$$

and with 0:

$$0 = 2 \cdot 3 = 3 \cdot 4 = 0 \cdot c = c \cdot 0, \quad \text{for } c \in \mathbb{Z}_6.$$

Thus, whenever 2 divides a product of two elements ab, then 2 divides a or b, and so is prime.

However, 2 is not irreducible: for example $2 = 2 \cdot 4$, but neither 2 nor 4 are units.

We also know $2 = 4 \cdot 5$. However, $5 \cdot 5 = 1$ and so 5 is a unit. This, therefore, does not help us to prove that 2 is not irreducible.

Exercise 2.2

(a) List the units, irreducible and prime elements in (i) $\mathbb{Z}_6$ and (ii) $\mathbb{Z}_8$.

(b) Is $4 + 3i$ irreducible in $\mathbb{Z}[i]$, the ring of Gaussian integers? (Do not spend too long on this before referring to the solution.)

In the solution to part (b) of the above exercise, we suggested that the **complex modulus** $|a + bi| = \sqrt{a^2 + b^2}$ might be used to find a factorisation of the complex number $a + bi$. This approach will be invaluable later in the chapter.

In Example 2.7, what was it about $\mathbb{Z}_6$ that made the number 2 reducible? First, notice that 2 is a zero divisor in $\mathbb{Z}_6$, since $2 \cdot 3 = 0$. Using the basic axioms for a ring, we can now write 2 as a product of two proper factors:

You may wish to remind yourself how zero divisor was defined in Definition 1.10 of Chapter 11.

$$\begin{aligned} 2 &= 0 + 2 \\ &\equiv (2 \cdot 3) + (2 \cdot 1) \\ &\equiv 2 \cdot (3 + 1) \equiv 2 \cdot 4 \,(\text{mod } 6). \end{aligned}$$

However, here we should add a word of caution: in Exercise 2.2(a)(ii), you should have found that in $\mathbb{Z}_8$ the number 2 is both prime and irreducible, even though it is also a zero divisor. Following the same argument as for $\mathbb{Z}_6$, we have

$$2 = 0 + 2 \equiv (2 \cdot 4) + (2 \cdot 1) \equiv 2 \cdot 5 \,(\text{mod } 8),$$

and we get the factorisation $2 = 2 \cdot 5$. However, 5 is a unit in $\mathbb{Z}_8$, and so it is not a proper factor.

Thus, in some rings, simply being a zero divisor might not be enough to give us a prime non-irreducible element. However, we can escape these complexities by restricting our attention to rings with no zero divisors.

Definition 2.8 *Integral domain*

A commutative ring is an **integral domain** if it has no zero divisors.

Equivalently, a ring R is an integral domain if for $a, b \in R$, whenever $a \cdot b = 0$ then either $a = 0$ or $b = 0$.

Example 2.9 *Some integral domains*

(a) As we should expect, $\mathbb{Z}$ is an integral domain.

(b) Any field F is an integral domain by Lemma 1.17 of Chapter 11. Therefore, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all integral domains.

(c) Any subring of a field is also an integral domain.

In fact, every integral domain is a subring of some field. If R is an integral domain, the smallest field F for which R is a subring of F is called the **field of fractions**. We will encounter these again in Book E.

Example 2.10 *Some rings that are not integral domains*

(a) When $n \in \mathbb{N}$ is composite, the ring $\mathbb{Z}_n$ contains zero divisors, as was shown in Exercise 1.8 of Chapter 11.

(b) The ring $\mathbb{Z} \times \mathbb{Z} = \{(a, b) : a, b \in \mathbb{Z}\}$ also contains zero divisors. This ring was considered in Exercise 1.6 of Chapter 11.

The key property of integral domains is the ability to 'cancel' terms, which gives us the following equivalent way to define an integral domain.

Lemma 2.11 *Cancellation in an integral domain*

Let R be a commutative ring. Then R is an integral domain if, and only if, for $a, b, c \in R$ with c non-zero, $ac = bc$ implies $a = b$.

Proof First, assume R is an integral domain. We will show that if $ac = bc$ for a non-zero element c then $a = b$. From $ac = bc$ we get $0 = ac - bc = (a - b)c$. Now, since R is an integral domain and $c \neq 0$, we must have $a - b = 0$ and therefore $a = b$.

Note that $-bc = (-b)c$ by Proposition 1.5 of Chapter 11.

Conversely, suppose R is not an integral domain. Then there exist non-zero $a, b \in R$ such that $ab = 0$. We now have $a \cdot 0 = 0 = ab$. However, $b \neq 0$, and so $a0 = ab$ does not imply $0 = b$. ■

Reassuringly, we now find that every prime element is irreducible in an integral domain.

Theorem 2.12

Let R be an integral domain and let p be a prime element of R. Then p is irreducible.

Proof Suppose $p = ab$ for some $a, b \in R$. We need to show that one of a or b is a unit. Since p is prime, and $p \mid ab$, either $p \mid a$ or $p \mid b$. Without loss of generality, assume that $p \mid a$. Then $a = pq$ for some $q \in R$ and so $p = ab = pqb$. Now $p = pqb$, and by Lemma 2.11, this implies that $1 = qb$ and so b is a unit. ■

Now that we are certain that restricting our attention to integral domains is a step in the right direction, we can list some useful properties of division, all of which follow directly from the definition.

Lemma 2.13 *Properties of division in an integral domain*

Let R be an integral domain and let a, b, c be non-zero elements in R.

(a) If $a \mid c$, then $a \mid (c + na)$ for any $n \in R$.

(b) If $a \mid b$ and $b \mid a$, then a is an associate of b.

(c) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(d) If $a \mid b$ and $a \mid c$, then $a \mid (mb + nc)$ for any $m, n \in R$.

(e) Every element of R divides 0.

(f) If u is a unit in R, then $u \mid n$ for every $n \in R$.

Exercise 2.3

Prove Lemma 2.13.

Exercise 2.4

Prove that if S is a subring of an integral domain R, then S is an integral domain.

2.3 Norms

In the last subsection we established that prime elements need not be irreducible in an arbitrary ring, but this is not the case in integral domains. In this subsection, we will begin to consider the question of whether irreducible elements are always prime.

First, let us consider the ring of integers $\mathbb{Z}$. Translating into the language of this chapter, Euclid's Lemma for prime factors (Theorem 1.4 of Book A, Chapter 2) tells us that an irreducible element in $\mathbb{Z}$ satisfies Definition 2.6. Coupling the above with Theorem 2.12, we find that in $\mathbb{Z}$, the definitions of 'prime' and 'irreducible' do in fact coincide.

Will prime and irreducible always be the same in an integral domain, or do we need more conditions to ensure this? To investigate this question, we first need some theory.

Definition 2.14 *Norm*

Let R be an integral domain. A **norm** is a function

$$N : R \setminus \{0\} \to \mathbb{N} \cup \{0\},$$

mapping the non-zero elements of R to the non-negative integers.

As we will see later, we need Euclidean norms to generalise the Division Algorithm.

N is a **Euclidean norm** if for non-zero $a, b \in R$ we have

(a) $N(a) \leq N(ab)$, and

(b) there exist $q, r \in R$ such that $a = bq + r$, and either $r = 0$ or $N(r) < N(b)$.

N is a **multiplicative norm** if for non-zero $a, b \in R$ we have $N(ab) = N(a)N(b)$.

The idea behind these definitions is that they will give us a 'measure' of the relative sizes of elements of R. Let us look at some examples.

Example 2.15 *Some norms*

(a) In the ring of integers $\mathbb{Z}$, the usual norm of an element $a \in \mathbb{Z} \setminus \{0\}$ is $|a|$, the absolute value of a.

It is multiplicative since $|ab| = |a| \cdot |b|$, and the Division Algorithm shows that it is a Euclidean norm.

(b) For a polynomial ring $F[x]$ over a field F, the degree of a non-zero polynomial is a norm, $N(f) = \deg(f)$.

It is not multiplicative, since $\deg(fg) = \deg(f) + \deg(g)$, but we do have $\deg(f) \leq \deg(fg)$ and so the Division Algorithm for polynomials (Theorem 3.1 of Chapter 11) shows that it is Euclidean.

(c) In the Gaussian integers $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$, the complex modulus squared, $N(a + bi) = a^2 + b^2$, is a norm. N is multiplicative since $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$, and so we have

$$\begin{aligned} N((a + bi)(c + di)) &= (ac - bd)^2 + (ad + bc)^2 \\ &= (a^2 + b^2)(c^2 + d^2) \\ &= N(a + bi)N(c + di). \end{aligned}$$

We discuss this norm further in Worked Exercise 3.4.

We can now return to the question of whether irreducible elements need always be prime in an integral domain. In our investigation, we will actually also encounter our first example of non-unique factorisation.

Consider the ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$. In this ring, we can write the element 6 in two different ways:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

This factorisation in two different ways *looks* like it should give us an example: assuming the element $2 \in \mathbb{Z}[\sqrt{-5}]$ is irreducible, then $2 \mid 6$, but it looks unlikely that 2 is going to divide either $(1 + \sqrt{-5})$ or $(1 - \sqrt{-5})$. If this really is true (and remember, we haven't proved anything yet), then this will show us that 2 is not prime.

Moreover, if the elements $3, (1 + \sqrt{-5})$ and $(1 - \sqrt{-5})$ are also all irreducible and none is an associate of another, then this will give us an example of non-unique factorisation.

What does it mean for an element to have 'unique factorisation'? We will define this more formally later, but it should mean that it can be uniquely expressed as a product of irreducible elements.

We have just made rather a lot of assumptions about elements in $\mathbb{Z}[\sqrt{-5}]$, and so now we need to prove these. The following exercise will lead you through the formal details – we will use a norm function to help us to prove that elements are irreducible.

Exercise 2.5 *Finding an irreducible element that is not prime*

Let R be the ring $\mathbb{Z}[\sqrt{-5}]$.

(a) Show that R is an integral domain.

(b) Define a norm $N : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ by $N(a + b\sqrt{-5}) = a^2 + 5b^2$. Show that N is a multiplicative norm.

(c) Show that if $r = a + b\sqrt{-5} \neq 0$ and $b \neq 0$ then $N(r) \geq 5$. What can you say about r if $N(r) < 5$?

(d) Why is 2 an irreducible element of R?
Hint: what is $N(2)$?

(e) Show that 2 does not divide $1 + \sqrt{-5}$ or $1 - \sqrt{-5}$, and hence show that 2 is not prime in R.

Exercise 2.6 *Proving non-unique factorisation*

As in Exercise 2.5, let R be the ring $\mathbb{Z}[\sqrt{-5}]$, and let N be the multiplicative norm $N(a+b\sqrt{-5}) = a^2+5b^2$.

(a) Show that if u is a unit of R then $N(u) = 1$, and deduce that the only units of R are $1, -1$.

(b) Show that 3 is irreducible in R.

(c) Show that $1+\sqrt{-5}$ and $1-\sqrt{-5}$ are irreducible elements of R.

(d) Hence factorise 6 as a product of irreducible elements in two different ways in R.

3 Euclidean domains

To explore uniqueness of factorisation further, this section uses the Euclidean norm from the previous section to identify a family of integral domains that have a division algorithm.

3.1 Rings with a division algorithm

We saw in the last section that there are rings (and, indeed, integral domains) where irreducible elements need not be prime, and this allowed us to construct an example where factorisation into irreducibles is not unique. In order to isolate the reasons why this can happen, let us compare what we know for general rings with what we know for $\mathbb{Z}$, where uniqueness of factorisation does hold.

Look back for a moment at Euclid's Lemma for prime factors, Theorem 1.4 in Book A, Chapter 2. Translating into our language, this says 'if $p \in \mathbb{Z}$ is irreducible, and p divides the product $a_1a_2\cdots a_n$, then p divides a_i for some i'. In other words, in $\mathbb{Z}$, if p is irreducible then it is prime – this is exactly what we want to see. How did we prove this, and can we mimic this proof in a more general setting?

Following the proof of Theorem 1.4, we find that we need to know about 'highest common factors' and be certain that these exist. In $\mathbb{Z}$ this was Proposition 4.7 of Book A, Chapter 1, and central to this proof was the Division Algorithm. This is what we need to generalise, and we have some experience of this already, through our study of polynomial division in Chapter 11.

For $\mathbb{Z}$, the Division Algorithm (Theorem 4.1 of Book A, Chapter 1) means that any integer a can be divided by any positive integer b to leave a 'small' remainder, that is, $a = bq + r$ where $r, q \in \mathbb{Z}$, and r is in the range $0 \leq r < b$. For polynomials with coefficients in a field, we defined a division algorithm where 'smaller' meant of lower degree (Theorem 3.1 of Chapter 11). To measure 'smallness' in arbitrary rings, we can use the ready-made concept of a norm, which leads us to the following definition.

Definition 3.1 *Euclidean domain*

Let R be an integral domain. If there exists a Euclidean norm $N : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ on R, then we say that R is a **Euclidean domain**.

Recall that a Euclidean norm satisfies two conditions. For non-zero $a, b \in R$:

(i) $N(a) \leq N(ab)$

(ii) there exist $q, r \in R$ such that $a = bq + r$ where $N(r) < N(b)$ or $r = 0$.

Example 3.2 *Some Euclidean domains*

(a) The ring of integers $\mathbb{Z}$ is a Euclidean domain with norm $N(a) = |a|$ for $a \in \mathbb{Z} \setminus \{0\}$. We observed in Example 2.15(a) that this norm is Euclidean, which follows from the results of Book A, Chapter 1.

(b) Any field F is a Euclidean domain. Since $F \setminus \{0\}$ consists entirely of units, for any $a, b \in F \setminus \{0\}$ we can take $q = b^{-1}a \in F$ so that $a = bq$. In other words, we can always take remainder 0, and therefore our norm function needs to satisfy only $N(a) \leq N(ab)$. Trivially, we may set $N(a) = 1$ for all $a \in F \setminus \{0\}$.

(c) The polynomial ring $F[x]$ where F is a field and $N(f) = \deg(f)$ for non-zero $f \in F[x]$ is a Euclidean domain. This was proved in Chapter 11.

Notice that it is possible for a ring to be a Euclidean domain with respect to more than one norm. For example, the ring of integers is also a Euclidean domain with respect to the norm $N(a) = a^2$.

From the definition of Euclidean norm, we immediately derive the following division algorithm.

Theorem 3.3 *Division Algorithm for Euclidean domains*

Let R be a Euclidean domain with Euclidean norm N and let $a, b \in R$ with b non-zero. Then there exist $q, r \in R$ such that $a = bq + r$ with $N(r) < N(b)$ or $r = 0$.

The following worked exercise gives an example of how to establish that a given function is a Euclidean norm and shows how to find elements q and r as required for the Division Algorithm.

Worked Exercise 3.4

Prove that the ring of Gaussian integers $\mathbb{Z}[i]$ is a Euclidean domain with $N(x+yi) = x^2 + y^2 = |x+yi|^2$.

Solution

First, $\mathbb{Z}[i]$ is an integral domain since it is a subring of $\mathbb{C}$, and so it remains to show that N is a Euclidean norm.

For non-zero $a, b \in \mathbb{Z}[i]$,

$$N(ab) = N(a)N(b) \geq N(a)$$

since $N(b) \geq 1$ for all non-zero $b \in \mathbb{Z}[i]$.

Finally, for $a, b \in \mathbb{Z}[i]$ and $b \neq 0$, we need to find $q, r \in \mathbb{Z}[i]$ such that $a = bq + r$ and $N(r) < N(b)$. We let q be a point in $\mathbb{Z}[i]$ that is nearest to the complex number $z = \frac{a}{b}$ (Figure 3.1). The point z lies inside a 1×1 square, and so the distance from z to the nearest corner of the square (that is, the point q) is at most $\frac{\sqrt{2}}{2}$. In other words, the distance from z to q, $|z-q|$, satisfies

$$|z-q| \leq \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} < 1.$$

Now, let $r = a - bq$. Then $r \in \mathbb{Z}[i]$, and

$$N(r) = |r|^2 = |a-bq|^2 = |b|^2|z-q|^2 < |b|^2 = N(b),$$

which proves that N is a Euclidean norm.

Note that q might not be unique, but we are really interested in the distance from z; so any closest point q will do.

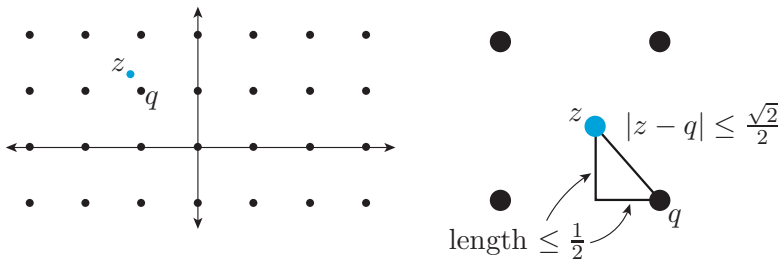


Figure 3.1 A nearest point $q \in \mathbb{Z}[i]$ to the point $z \in \mathbb{C}$

Exercise 3.1

(a) Following the approach in Worked Exercise 3.4, prove that $\mathbb{Z}[\sqrt{-2}]$ is a Euclidean domain, with norm $N(a+b\sqrt{-2}) = a^2+2b^2 = |a+b\sqrt{2}i|^2$.

(b) Explain why a similar argument cannot be used for $\mathbb{Z}[\sqrt{-5}]$ with the norm $N(a+b\sqrt{-5}) = a^2+5b^2$.

(c) By considering the elements 2 and $1+\sqrt{-5}$, show that $\mathbb{Z}[\sqrt{-5}]$ is *not* a Euclidean domain with respect to the norm $N(a+b\sqrt{-5}) = a^2+5b^2$.

Before we use our newly formed Division Algorithm to advance our story, let us pause briefly to record some basic properties of Euclidean norms.

Lemma 3.5

Let R be a Euclidean domain with norm N. Then:

(a) $N(1)$ is the smallest value taken by the function N over all non-zero elements of R.

(b) $N(u) = N(1)$ if, and only if, u is a unit in R.

(c) If a, b are associates in R then $N(a) = N(b)$.

(d) For any non-zero $a \in R$, if $b \neq 0$ is not a unit then $N(a) < N(ab)$.

Proof

(a) For any non-zero element $b \in R$, we have $N(b \cdot 1) \geq N(1)$ from Definition 2.14 and so $N(b) \geq N(1)$.

(b) Suppose that u is a unit in R. Then $N(u) \leq N(uu^{-1}) = N(1)$ $\leq N(u \times 1) = N(u)$, and so $N(u) = N(1)$.

Conversely, suppose $u \neq 0 \in R$ and $N(u) = N(1)$. Then there exist $q, r \in R$ such that $1 = uq + r$ with either $N(r) < N(u) = N(1)$ or $r = 0$. Since we cannot have $N(r) < N(1)$ by part (a), it follows that $r = 0$ and $1 = uq$. In other words, u is a unit in R.

(c) Since a and b are associates, there is a unit $u \in R$ so that $au = b$. Then $N(b) = N(au) \geq N(a)$ and $N(a) = N(bu^{-1}) \geq N(b)$. Hence $N(a) = N(b)$.

(d) Suppose that $a, b \in R$ are non-zero and that b is not a unit. Since N is Euclidean, we have $N(a) \leq N(ab)$. If $N(a) = N(ab)$, we want to conclude that b is a unit, which will be a contradiction. Since a and ab are non-zero, we can find q, r such that

$$a = (ab)q + r,$$

where either $r = 0$ or $N(r) < N(ab) = N(a)$. Rearranging, we have $r = a(1-bq)$. We cannot have $r \neq 0$, as this would imply $N(r) = N(a(1-bq)) \geq N(a)$, which is impossible. Therefore $a(1-bq) = 0$, and since $a \neq 0$ we must have $1 - bq = 0$. This means that $bq = 1$, and so b is a unit, as required. ■

The above lemma gives us a method of computing units, as the next exercise shows.

Exercise 3.2

(a) Prove that the only units in $\mathbb{Z}[i]$ are $\pm 1, \pm i$.

(b) Compute the units in the Euclidean domain $\mathbb{Z}[\sqrt{-2}]$, with norm $N(a + b\sqrt{-2}) = a^2 + 2b^2$.

3.2 Highest common factors

We begin this subsection with a direct generalisation of the definition of highest common factor, to apply to an integral domain.

Definition 3.6 *Highest common factor*

Let a and b be elements of an integral domain R. An element h is said to be a **highest common factor** (**hcf**) of a and b if it satisfies the following:

(a) $h \mid a$ and $h \mid b$

(b) if $d \mid a$ and $d \mid b$, then $d \mid h$.

Highest common factors are also known as **greatest common divisors**, denoted by **gcd**.

Condition (a) says that a highest common factor *is* a common factor of elements, while condition (b) says that it is *maximal* amongst the common factors. Note that we can talk only about *a* (rather than *the*) highest common factor because if $h = \text{hcf}(a, b)$, then any associate $u \cdot h$ (where u is a unit) also satisfies conditions (a) and (b). In our definition of highest common factor for the integers we specified that we would take the positive number to ensure uniqueness, and for polynomial rings $F[x]$ over fields we fixed the highest common factor to be the unique monic polynomial, but in this more general setting we cannot meaningfully enforce such a rule, and so we have to accept non-uniqueness.

As well as non-uniqueness, there also remains the question of whether hcfs must exist. First, as we did in $\mathbb{Z}$ and $F[x]$, we adopt the convention that $\text{hcf}(0, 0)$ never exists, and so we should consider $\text{hcf}(a, b)$ only where at least one of a or b is non-zero. In fact, considering only integral domains, there are still instances where hcfs do not exist.

Worked Exercise 3.7

Consider the ring $\mathbb{Z}[\sqrt{-5}]$. Prove that the elements 6 and $2(1+\sqrt{-5})$ do not have a highest common factor.

Solution

We will make use of the multiplicative norm $N(a+b\sqrt{-5}) = a^2+5b^2$ (introduced in Exercise 2.5). First, $N(6) = 36$ and $N(2(1+\sqrt{-5})) = 4+20 = 24$, and so if $h = \text{hcf}(6, 2(1+\sqrt{-5}))$ exists then $N(h)$ must divide 24 and 36. Thus $N(h)$ must be equal to one of 1, 2, 3, 4, 6 or 12.

As we saw in Exercise 2.5, the element 6 can be written as the product of irreducibles in two different ways:

$$6 = 2\cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5}).$$

Consequently, 6 and $2(1+\sqrt{-5})$ have both 2 and $1+\sqrt{-5}$ as common factors, and so both of these must divide h. Now $N(2) = 4$ and $N(1+\sqrt{-5}) = 6$, and so 4 and 6 must both divide $N(h)$, which implies that the only possibility is that $N(h) = 12$.

Writing $h = a+b\sqrt{-5}$, we have $N(h) = a^2+5b^2 = 12$. From this, the only options are $b = 0$ or 1, for otherwise $a^2+5b^2 \geq 20 > 12$. When $b = 0$ we have $a^2 = 12$, while if $b = 1$ then $a^2 = 12-5 = 7$; neither of these has solutions in $\mathbb{Z}$. Therefore h cannot exist.

When we proved the existence of hcfs in the integers (and in polynomial rings over fields), we repeatedly applied the Division Algorithm until the remainder was 0. We were guaranteed that this repeated application of the Division Algorithm would stop, because in the equation $a = bq+r$ we ensured $0 \leq r < b$. For Euclidean domains, the equivalent condition is 'either $N(r) < N(b)$ or $r = 0$', and this gives us the tool we need to define the **Euclidean Algorithm** (and consequently to prove the existence of hcfs) for arbitrary Euclidean domains.

Theorem 3.8

Let R be a Euclidean domain with norm N, and let $a, b \in R$ with b non-zero. Then a and b have a highest common factor. If $h = \text{hcf}(a,b)$, then there are elements $s, t \in R$ so that $h = sa+tb$.

Proof We begin by defining $S = \{xa + yb : x, y \in R\}$, and notice that S is non-empty because $a \in S$ and $b \in S$, as shown by $a = 1 \cdot a + 0 \cdot b$ and $b = 0 \cdot a + 1 \cdot b$.

Now, by the Well-Ordering Principle from Book A, Chapter 1, we choose non-zero $h \in S$ such that $N(h)$ is as small as possible. Since $h \in S$, we know that $h = sa + tb$ for some $s, t \in R$.

It remains to show that h is indeed a highest common factor of a and b. First, if $d \mid a$ and $d \mid b$, then clearly $d \mid xa + yb$ for any $x, y \in R$, and so $d \mid h$. Since R is a Euclidean domain, by Definition 3.1 we have $a = qh + r$ for some $q, r \in R$, with either $r = 0$ or $N(r) < N(h)$. Now, if $r \neq 0$ then

$$r = a - qh = a - q(sa + tb) = (1 - qs)a - (qt)b \in S.$$

We now have a contradiction because $r \in S$ and $N(r) < N(h)$, but h was chosen to minimise $N(h)$ over all non-zero elements in S. Therefore $r = 0$, $a = qh$, and consequently $h \mid a$. A similar argument can be applied to show that $h \mid b$. ■

In the solution to Exercise 3.1(c), we showed that $\mathbb{Z}[\sqrt{-5}]$ was not a Euclidean domain with respect to the norm $N(a + b\sqrt{-5}) = a^2 + 5b^2$. Could it be a Euclidean domain with respect to some other norm? Theorem 3.8 now gives us a way of showing that the answer to this question is no.

Worked Exercise 3.9

Show that $\mathbb{Z}[\sqrt{-5}]$ is not a Euclidean domain with respect to any norm.

Solution

We proved (in Worked Exercise 3.7) that the elements 6 and $2(1 + \sqrt{-5})$ have no highest common factor in $\mathbb{Z}[\sqrt{-5}]$. Therefore $\mathbb{Z}[\sqrt{-5}]$ cannot be a Euclidean domain with respect to any norm, by Theorem 3.8.

Note that even though we specified a norm to prove that 6 and $2(1 + \sqrt{-5})$ have no hcf, the hcf itself is defined without reference to any norm.

Theorem 3.8 tells us not only that highest common factors exist in Euclidean domains, but that $\text{hcf}(a, b)$ must be expressible as a linear combination of a and b. This additional requirement gives us another method to prove that some rings are not Euclidean domains, as the next exercise shows.

Exercise 3.3

(a) Show that $\mathrm{hcf}(2, x) = 1$ in $\mathbb{Z}[x]$.

(b) Hence, or otherwise, conclude that $\mathbb{Z}[x]$ is not a Euclidean domain.

(c) Find polynomials $s, t \in \mathbb{Q}[x]$ such that $2s(x) + xt(x) = 1$ in $\mathbb{Q}[x]$.

The existence of highest common factors in Euclidean domains now gives us a route to what we have been working towards in this section: situations where irreducible and prime mean the same thing. First, we need a preliminary result.

Lemma 3.10

Suppose that p is an irreducible element of a Euclidean domain R, that a is a non-zero element of R and that p does not divide a. Then there exist $s, t \in R$ with $ps + at = 1$.

Proof Let b be a highest common factor of a and p. Then by Theorem 3.8 there exist $s', t' \in R$ with $b = s'a + t'p$. Now since b divides p and p is irreducible, either $b = up$ where u is a unit, or b is a unit. If $b = up$ then, since b divides a, p also divides a. Thus b is a unit and so $1 = b^{-1}b = b^{-1}s'a + b^{-1}t'p$ and the result follows with $s = b^{-1}s'$ and $t = b^{-1}t'$. ■

We are now ready for the following result.

Proposition 3.11

Let R be a Euclidean domain with norm N, and let p be an irreducible element in R. Then p is prime in R.

Proof Let $a, b \in R$ be such that p divides ab. If p divides a, then we have finished; so let us suppose that p does not divide a. By Lemma 3.10, there exist $s, t \in R$ such that

$$ps + at = 1.$$

Multiplying this equation by b, we obtain

$$pbs + abt = b.$$

Now, since $p \mid ab$, and p must divide itself, we can apply Lemma 2.13(d) to show that p divides $pbs + abt$. Since $b = pbs + abt$, we finally conclude that p must divide b. Hence p is prime. ■

Recall that to prove p is prime, we need to show that whenever $p \mid ab$, then $p \mid a$ or $p \mid b$.

3.3 Prime numbers expressed as the sum of two squares

Before we continue our exploration of unique factorisation, let us pause for a moment to present a number theoretic result that follows from our study of Euclidean domains.

Another problem that attracted Fermat's attention concerned ways of expressing positive integers as sums of squares. For example, the integers from 1 to 9 can be written as sums of squares as follows.

$$1 = 1^2$$
$$2 = 1^2 + 1^2$$
$$3 = 1^2 + 1^2 + 1^2$$
$$4 = 2^2$$
$$5 = 2^2 + 1^2$$
$$6 = 2^2 + 1^2 + 1^2$$
$$7 = 2^2 + 1^2 + 1^2 + 1^2$$
$$8 = 2^2 + 2^2$$
$$9 = 3^2$$

The expressions given are certainly not unique; for example, we could write

$$4 = 1^2 + 1^2 + 1^2 + 1^2 \quad \text{or} \quad 9 = 2^2 + 2^2 + 1^2.$$

But notice that the given expression for 7 as a sum of squares is the best we can do in the sense that 7 cannot be expressed as a sum of fewer than four squares. If you try continuing the list that we have started, with the aim of expressing each integer as a sum of as few squares as possible, you should find that each requires no more than four squares.

Amongst other questions that these observations might suggest are the following.

We accept 0 as a square so that, for example, $4 = 2^2 + 0^2$ is a legitimate way of expressing 4 as a sum of two squares. However, we ignore negative integers since $(-a)^2 = a^2$.

- Which integers can be expressed as a sum of two squares?
- Which integers can be expressed as a sum of two squares in a unique way?
- Which integers can be expressed as a sum of three squares?
- Can all positive integers be expressed as a sum of four squares?

In this subsection, we begin an investigation into the two-squares problem. The following identity, which is easily established, is going to be crucial in what follows.

Important identity for two squares

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

This identity should look familiar: we used it when proving that the complex modulus, $|a + bi|^2 = a^2 + b^2$, is a multiplicative norm in Example 2.15(c) of this chapter.

This identity is of theoretical importance because it tells us that if two positive integers m and n can each be written as a sum of two squares then

so too can their product mn. So in order to show that a given integer can be written as a sum of two squares, it is sufficient to show that each prime in its decomposition can be expressed in this way.

But the identity also has a practical use, as illustrated by the following example.

Worked Exercise 3.12

Express 4420 as a sum of two squares.

Solution

$$4420 = 2^2 \times 5 \times 13 \times 17 = (2^2+0^2)(2^2+1^2)(3^2+2^2)(4^2+1^2)$$

Using the identity three times:

$$\begin{aligned} 4420 &= (4^2+2^2)(3^2+2^2)(4^2+1^2) \\ &= (16^2+2^2)(4^2+1^2) \\ &= 66^2+8^2. \end{aligned}$$

This answer is not unique. It turns out that there are four ways of writing 4420 as a sum of two squares. The other ways can all be obtained from the identity applied to different terms. We could change the order of the bracketed terms, or switch the two terms within a pair of brackets, or make use of negative values for a, b, c or d. Each of these variations is illustrated below, where we discover the other three solutions.

$$\begin{aligned} 4420 &= (4^2+2^2)(4^2+1^2)(2^2+3^2) \\ &= (4^2+2^2)(11^2+10^2) = 64^2+18^2 \\ 4420 &= (4^2+2^2)(11^2+(-10)^2) = 24^2+(-62)^2 = 62^2+24^2 \\ 4420 &= ((4^2+1^2)(2^2+1^2))((2^2+0^2)(2^2+3^2)) \\ &= (9^2+2^2)(4^2+6^2) = 48^2+46^2 \end{aligned}$$

Exercise 3.4

Find three ways of expressing 325 as a sum of two squares.

Turning to the question of whether a positive integer n can be expressed as a sum of two squares, our important identity guides us to look at the primes occurring in its decomposition. Now primes can be classified into three types:

- the even prime $2 = 1^2 + 1^2$, which is a sum of two squares
- odd primes of the form $4k + 1$
- odd primes of the form $4k + 3$.

The third category presents little problem. As a square is congruent modulo 4 to either 0 or 1, a sum of two squares is congruent modulo 4 to one of 0, 1 or 2. So no prime of the form $4k + 3$ can be expressed as a sum of two squares.

That leaves just the second category. We will prove that all primes of the form $4k + 1$ can be written as a sum of two squares. The original proof, due to Fermat, used the infinite descent method that we encountered in Section 1, albeit in a slightly different way. The (shorter) proof that we present here makes use of the fact that the ring of Gaussian integers, $\mathbb{Z}[i]$, is a Euclidean domain with respect to the norm $N(a + bi) = |a + bi|^2 = a^2 + b^2$.

Theorem 3.13

A prime number p can be expressed as the sum of two squares if, and only if, it is either even, or odd of the form $4k + 1$ for some integer k.

Proof First, if $p = u^2 + v^2$ is an odd number, then (by the comments made before the statement of the theorem) $p = 4k + 1$, which, together with noting that $2 = 1 + 1$ is the sum of two squares, proves the 'only if' part.

For the other direction, we first of all recall Theorem 2.2 from Chapter 10: since $p \equiv 1 \pmod 4$, we have the Legendre symbol $(-1/p) = 1$, that is, -1 is a square modulo p. This means that there must exist $a \in \mathbb{Z}_p$ such that $a^2 \equiv -1 \pmod p$. Thus, p divides $a^2 + 1$.

We now start working in $\mathbb{Z}[i]$: first, note that $a^2 + 1 = (a + i)(a - i)$, and so if p is prime in $\mathbb{Z}[i]$ then p must divide either $a + i$ or $a - i$. This, however, is impossible as p does not divide the imaginary part of either $a + i$ or $a - i$. Therefore, p is not prime and hence, since $\mathbb{Z}[i]$ is a Euclidean domain, p is not irreducible in $\mathbb{Z}[i]$ by Proposition 3.11. Therefore we can write $p = (a + ib)(c + id)$ as a product of two non-units in $\mathbb{Z}[i]$.

If $p \mid a + i$, then
$$a + i = p \cdot (b + ci) = pb + (pc)i$$
for some b, c in $\mathbb{Z}$. But then $pc = 1$, which in $\mathbb{Z}$ is impossible since $p > 1$.

Taking norms, we have $p^2 = (a^2 + b^2)(c^2 + d^2)$, and as $a + ib$ and $c + id$ are not units, both $a^2 + b^2$ and $c^2 + d^2$ must be greater than 1. It follows from the Fundamental Theorem of Arithmetic that $p = a^2 + b^2$ is the sum of two squares. ■

Knowing that a prime $p = 4k + 1$ can be expressed as $x^2 + y^2$, mathematicians understandably took up the challenge of finding ways of constructing the integers x and y in terms of the prime p. Several such constructions have emerged, but none of them is particularly easy to employ.

For practical purposes, the method of exhaustion offers as good a method as any for expressing moderately sized primes as a sum of two squares. For instance, consider the prime 277. First note that if $277 = x^2 + y^2$ then one of x^2 or y^2 is less than $\frac{277}{2}$, while the other is greater than $\frac{277}{2}$. By virtue of Theorem 3.13 we need to check

$$277 - 1^2, \quad 277 - 2^2, \quad 277 - 3^2, \quad \ldots$$

only until we find a square, in the certain knowledge that one will turn up by the time we reach $277 - 12^2$, since $12^2 > \frac{277}{2}$. In fact $277 - 9^2 = 14^2$.

To represent a composite number as a sum of two squares, we can combine the suggested exhaustive approach with use of the important identity.

Exercise 3.5

Express $5321 = 17 \times 313$ as a sum of two squares.

4 Unique factorisation domains

This section takes a direct approach to unique factorisation by naming integral domains that have this property 'unique factorisation domains'.

4.1 Prime and irreducible in unique factorisation domains

In Exercise 2.5 of this chapter, we found that there are irreducible elements in $\mathbb{Z}[\sqrt{-5}]$ that are not prime. In Exercise 2.6, we used the irreducible non-prime elements from Exercise 2.5 to establish that the element 6 of $\mathbb{Z}[\sqrt{-5}]$ can be written as a product of irreducibles in more than one way:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

We now explore rings where this is not the case for any element. We begin by clarifying exactly what we mean when we say that a ring has unique factorisation. The complications in this definition arise solely to handle units. For example, writing $6 = 2 \cdot 3 = (-2) \cdot (-3)$ should not mean that 6 does not factorise uniquely in $\mathbb{Z}$.

Definition 4.1 *Unique factorisation domain*

An integral domain R is a **unique factorisation domain** (UFD) if

(a) every non-zero non-unit element of R is expressible as a product of irreducible elements of R, and

(b) if $a \in R$ can be written as a product of irreducibles in two different ways,

$$a = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n,$$

then $m = n$ and the p_i, q_j pair off as associates.

Condition (a) of this definition tells us that a factorisation into irreducibles must *exist*, while part (b) ensures the *uniqueness*, up to associates. Our first example of a UFD is the ring of integers $\mathbb{Z}$: this, of course, is exactly what the Fundamental Theorem of Arithmetic tells us. On the other hand, we showed in Exercise 2.6 that the ring $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

We are now, at last, in a position to complete the story about the significance of prime and irreducible elements. The next theorem gives us a different characterisation of UFDs. As we have found in the examples we have looked at so far, irreducible and prime coincide precisely when elements factorise uniquely. Of course, we still need to be certain that we can always find *some* factorisation into irreducibles (condition (a) of Definition 4.1).

Theorem 4.2

Let R be an integral domain in which every non-zero non-unit element has a factorisation into irreducible elements. Then R is a UFD if, and only if, every irreducible element in R is prime.

Proof Suppose first that every irreducible element in R is prime. We will use a 'minimal counter-example' proof. Suppose that there exists $a \in R$ with two factorisations into irreducibles,

$$a = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n,$$

where the p_i and q_j do not pair off as associates. By swapping the two factorisations if necessary we can ensure $m \leq n$, and by the Well-Ordering Principle we choose such an a so that m is as small as possible. Note that if $m = 1$, then a itself is irreducible and so cannot be a product of more than one irreducible. This would imply that $a = p_1 = q_1$, which is a contradiction, and therefore $m > 1$.

This is related to mathematical induction. Supposing the statement is false, we take a smallest counter-example, and use the fact that anything smaller cannot be a counter-example to derive a contradiction.

Now, since p_m is prime and divides $a = q_1 q_2 \cdots q_n$, p_m must divide q_j for some j. This means that $q_j = u p_m$ for some $u \in R$, and since q_j is irreducible, u must be a unit. Now, however, if we write $b = p_1 p_2 \cdots p_{m-1}$, then

$$b p_m = a = q_1 q_2 \cdots q_n = q_1 \cdots q_{j-1} (u p_m) q_{j+1} \cdots q_n.$$

Cancelling p_m from either side (as we may since R is an integral domain), we have

$$b = p_1 \cdots p_{m-1} = q_1 q_2 \cdots q_{j-1}(u q_{j+1}) \cdots q_n.$$

We chose m to be as small as possible, and therefore b must have unique factorisation. This means that $m - 1 = n - 1$ (and so $m = n$), and the p_i $(1 \leq i \leq m-1)$ pair off with the remaining q_k as associates. However, we also paired off p_m with q_j, which means we have paired off all the p_i with the q_k, and this is a contradiction.

Conversely, suppose that R is a UFD. Consider an irreducible p for which $p \mid ab$ for some $a, b \in R$. That is, there is an $h \in R$ such that $ph = ab$. Then by unique factorisation, $a = p_1 \cdots p_n$, $b = q_1 \cdots q_m$ and $h = h_1 \cdots h_r$ where $p_1, \ldots, p_n, q_1, \ldots, q_m, h_1, \ldots, h_r$ are irreducible elements in R. This gives

$$ph_1 \cdots h_r = p_1 \cdots p_n q_1 \cdots q_m.$$

By the uniqueness of the factorisation, either p is an associate of p_i for some $i \in \{1, \ldots, n\}$ or p is an associate of q_j for some $j \in \{1, \ldots, m\}$. Hence either p divides a or p divides b. ■

Note that as well as providing a characterisation of UFDs, Theorem 4.2 also shows us that the terms 'irreducible' and 'prime' must coincide in a UFD. Although just a direct consequence of the above theorem and Definition 4.1, we record it here as a corollary.

Corollary 4.3

In a unique factorisation domain, every irreducible element is prime.

An important application of this corollary comes via its contrapositive: if a ring contains an irreducible element that is not prime, then it cannot be a UFD.

For example, consider again the ring $\mathbb{Z}[\sqrt{-5}]$. We know from Exercise 2.6 that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD since the element 6 can be written as a product of irreducibles in two different ways, $6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. However, by using Corollary 4.3 we could have deduced this at the end of Exercise 2.5: the element 2 is irreducible but not prime, and so $\mathbb{Z}[\sqrt{-5}]$ certainly cannot be a UFD. Of course, it is more informative to find an explicit example of non-unique factorisation; so the process in Exercise 2.6 is still a valuable technique to understand.

The next exercise asks you to prove the existence of least common multiples in UFDs. Before you can do so, we need to generalise the notion of least common multiple to integral domains.

Definition 4.4 *Least common multiple*

Let a and b be elements of an integral domain R. An element ℓ is said to be a **least common multiple** (**lcm**) of a and b if it satisfies the following:

(a) a divides ℓ and b divides ℓ

(b) for any $h \in R$ such that a divides h and b divides h, then ℓ also divides h.

Exercise 4.1

Prove that if R is a UFD, $\text{lcm}(a, b)$ exists for any non-zero $a, b \in R$.

4.2 UFDs and Euclidean domains

Since 'irreducible' and 'prime' mean the same thing in both Euclidean domains and UFDs, we might naturally wonder whether there is a relationship between these two concepts. In particular, we consider two questions: do all Euclidean domains have unique factorisation, and conversely, is every UFD a Euclidean domain? In light of Theorem 4.2, we can answer the first of these questions straight away.

Theorem 4.5

Every Euclidean domain is a unique factorisation domain.

Proof By Proposition 3.11 and Theorem 4.2, it suffices to show that each non-zero non-unit element of a Euclidean domain has some factorisation into irreducibles. Let R be a Euclidean ring with norm N, and let $a \in R$. We will again use a minimal counter-example proof. Let

$$S = \{x \in R : x \text{ is not zero and not a unit and cannot be written as a product of irreducibles}\}.$$

We want to show that S is empty. So, to derive a contradiction, let us suppose that it is non-empty. Using the Well-Ordering Principle, we choose $x \in S$ such that $N(x)$ is as small as possible.

Now since x cannot be written as a product of irreducibles, it cannot itself be irreducible. Hence there are elements $a, b \in R$, neither of which is a unit, with $ab = x$. Then, using Lemma 3.5(d),

$$N(a) < N(ab) = N(x) \text{ and } N(b) < N(ab) = N(x).$$

Hence neither a nor b belongs to S and so both can be written as products of irreducibles. Thus x can be written as a product of irreducibles, contradicting our assumption that S is non-empty. ■

In Worked Exercise 3.9, we proved that $\mathbb{Z}[\sqrt{-5}]$ was not a Euclidean domain with respect to any norm, because there are pairs of numbers that did not have a highest common factor. We now have another route to prove this, by using the contrapositive statement to Theorem 4.5: a ring that is not a UFD cannot be a Euclidean domain. Since we know that $\mathbb{Z}[\sqrt{-5}]$ is not a UFD (either by Exercise 2.6 or by the comments following Corollary 4.3), we can immediately deduce that $\mathbb{Z}[\sqrt{-5}]$ is not a Euclidean domain.

Recall that, by Corollary 4.3, in order to show non-uniqueness of factorisation it suffices to find an element that is irreducible but not prime. To do this, we follow the process used in Exercise 2.5, using a multiplicative norm to prove irreducibility and indivisibility. The following exercise invites you to try this.

Exercise 4.2

Show that $\mathbb{Z}[\sqrt{-6}]$ is not a Euclidean domain.

Hint: using the norm $N(a+b\sqrt{-6}) = a^2+6b^2$, consider the element 2, and factorisations of the element 6.

The second question that we asked at the start of this subsection was whether every UFD is a Euclidean domain. The answer, perhaps surprisingly, is 'no'.

Example 4.6 *A UFD that is not a Euclidean domain*

In Exercise 3.3 we established that $\mathbb{Z}[x]$ is not a Euclidean domain, by demonstrating that there cannot exist integer polynomials $s(x)$ and $t(x)$ for which $2s(x)+xt(x) = \text{hcf}(2,x) = 1$.

However, we will prove shortly that $\mathbb{Z}[x]$ is a UFD.

Note that $\mathbb{Z}[x]$ fails to be a Euclidean domain, even though hcfs exist in it, because we cannot find suitable polynomials $s(x)$ and $t(x)$. In fact, in a unique factorisation domain, hcfs *must* exist.

Proposition 4.7

Let R be a unique factorisation domain and $a, b \in R$, not both of which are zero. Then $\text{hcf}(a,b)$ exists in R.

Proof If one of a, b is zero, then the non-zero one will be the highest common factor. If both are non-zero, then a and b factorise uniquely into a product of irreducibles, and so the highest common factor is the product of irreducible factors that occur in both factorisations. ■

On the other hand, rings such as $\mathbb{Z}[\sqrt{-5}]$ do not have unique factorisation, and in Worked Exercise 3.7 we used this to find a pair of elements that does not have a highest common factor.

We now turn to proving that $\mathbb{Z}[x]$ is a UFD. To do this, we will use the characterisation of UFDs given by Theorem 4.2, which involves two steps: the first is to show that every polynomial in $\mathbb{Z}[x]$ can be factorised into a product of irreducibles, and the second is that every irreducible polynomial in $\mathbb{Z}[x]$ is prime. In Chapter 17, Book E, we will use the same approach to prove a more general result, namely that $R[x]$ is a UFD whenever R is.

You may find it helpful to re-read Subsections 4.2 and 4.3 of Chapter 11 before continuing.

As we work through the proof, we will need to recall several definitions and results from Chapter 11. Recall that for a polynomial $f \in \mathbb{Z}[x]$, the **content**, c_f, of f is the highest common factor of all of its coefficients, and that if $c_f = 1$ then we say that f is **primitive**.

Lemma 4.8

Every polynomial in $\mathbb{Z}[x]$ has a factorisation into irreducible polynomials in $\mathbb{Z}[x]$.

Proof By the Fundamental Theorem of Arithmetic, the content c_f of a polynomial $f \in \mathbb{Z}[x]$ has a unique factorisation $c_f = p_1 p_2 \cdots p_s$ in $\mathbb{Z}$. Since each p_i is irreducible in $\mathbb{Z}$, it is also irreducible in $\mathbb{Z}[x]$, and so we can write $f = c_f f^*$ where f^* is primitive, and c_f has a unique factorisation into irreducibles. Consequently, it suffices to prove the lemma only for primitive polynomials in $\mathbb{Z}[x]$. Define

Remember that the irreducible elements of $\mathbb{Z}$ are the prime numbers.

$$S = \{f \in \mathbb{Z}[x] : f(x) \text{ is primitive, non-zero, is not } \pm 1 \text{ and cannot be written as a product of irreducibles}\}.$$

We need to show that S is empty. Suppose to the contrary that S is non-empty, and by the Well-Ordering Principle take a minimal counter-example. That is, take $f(x) \in S$ such that $\deg(f)$ is as small as possible.

Notice the similarity between this proof and the proof of Theorem 4.5.

Since f cannot be written as a product of irreducibles, it cannot itself be irreducible, and so it must have a proper factorisation $f = ab$ for some polynomials $a, b \in \mathbb{Z}[x]$. Since f is primitive, Lemma 4.12 of Chapter 11 tells us that both a and b are also primitive, and so in order for $f = ab$ to be a proper factorisation, we must have $\deg(a) \geq 1$ and $\deg(b) \geq 1$.

From $\deg(f) = \deg(ab) = \deg(a) + \deg(b)$ we know that $\deg(a) < \deg(f)$ and $\deg(b) < \deg(f)$. Since f was a polynomial of minimal degree in S, each of a and b is either irreducible, or can be written as a product of irreducibles. In every case, however, we reach a contradiction: $f = ab$ can be written as a product of irreducibles, even though $f \in S$. Consequently S is empty, and all primitive polynomials in $\mathbb{Z}[x]$ can be written as a product of irreducibles. ■

The second step needed to complete the proof that $\mathbb{Z}[x]$ is a UFD is to show that every irreducible element of $\mathbb{Z}[x]$ is prime.

Proposition 4.9

Let f be an irreducible polynomial in $\mathbb{Z}[x]$. Then f is prime in $\mathbb{Z}[x]$.

We divide the proof into two cases. The first is where we are given a polynomial $f \in \mathbb{Z}[x]$ for which $\deg(f) = 0$. In this case f is an integer, and since irreducible and prime mean the same thing in $\mathbb{Z}$, f is prime in $\mathbb{Z}$. The following lemma completes this case.

Lemma 4.10

Let $p \in \mathbb{Z}$. If p is prime in $\mathbb{Z}$, then it is prime in $\mathbb{Z}[x]$.

Proof Following a minimal counter-example argument, suppose that p is prime in $\mathbb{Z}$ but not in $\mathbb{Z}[x]$. Then there are polynomials $f, g \in \mathbb{Z}[x]$ such that p divides fg, but p divides neither f nor g. If $\deg(f) = m$ and $\deg(g) = n$, then $\deg(fg) = m + n$, and by the Well-Ordering Principle we can choose f and g so that $m + n$ is as small as possible.

Now from the fact that p divides fg it follows that p must divide every coefficient of fg. Denote the leading coefficients of f and g by a_m and b_n, respectively, so that the leading coefficient of fg is $a_m b_n$. Since p is prime in $\mathbb{Z}$, p must divide at least one of a_m or b_n, and without loss of generality suppose that p divides a_m. Note that $m > 0$ because otherwise p would divide f. Next, p divides

$$f(x)g(x) - a_m x^m g(x) = (f(x) - a_m x^m) \cdot g(x),$$

and $\deg(f(x) - a_m x^m) < m$. So, by our assumption that $m + n$ was as small as possible, p must divide $f(x) - a_m x^m$ since it does not divide g. However, p divides a_m, and so it divides $f(x) - a_m x^m + a_m x^m = f(x)$, which is a contradiction. ■

The other case, where $f \in \mathbb{Z}[x]$ and $\deg(f) \geq 1$, requires a different approach. The key step is to consider a polynomial in $\mathbb{Z}[x]$ as a polynomial in $\mathbb{Q}[x]$, and use the fact that irreducible and prime mean the same thing in $\mathbb{Q}[x]$ (since it is a Euclidean domain – see Example 3.2(c) of this chapter). This approach now enables us to give the formal proof of Proposition 4.9.

Proof of Proposition 4.9 Following the pattern laid out in the discussion above, we identify two cases. First, if $\deg(f) = 0$ then $f \in \mathbb{Z}$ is irreducible in $\mathbb{Z}$, and hence (since $\mathbb{Z}$ is a Euclidean domain) f is also prime in $\mathbb{Z}$. By Lemma 4.10, f is prime in $\mathbb{Z}[x]$.

Now suppose that $\deg(f) \geq 1$. Take $g, h \in \mathbb{Z}[x]$ for which f divides gh. To prove that f is prime in $\mathbb{Z}[x]$, we need to show that f divides one of g or h.

Since f is irreducible in $\mathbb{Z}[x]$ and $\deg(f) \geq 1$, f is primitive (otherwise $f = c_f f^*$ is a proper factorisation). Therefore, by Gauss's Lemma (Theorem 4.13 of Chapter 11), f is irreducible in $\mathbb{Q}[x]$. Since $\mathbb{Q}[x]$ is a

Euclidean domain, f is prime in $\mathbb{Q}[x]$, which means that in $\mathbb{Q}[x]$ either f divides g or f divides h, and without loss of generality we assume that f divides g. Therefore there exists $s \in \mathbb{Q}[x]$ such that $fs = g$. We want to show that the coefficients of $s(x)$ are all integers, so that f divides g in $\mathbb{Z}[x]$. Writing $s(x) = \frac{p}{q}t(x)$ where $t(x) \in \mathbb{Z}[x]$ is primitive and $\frac{p}{q} \in \mathbb{Q}$ is written in its lowest terms, we have, in $\mathbb{Q}[x]$,

The comments before Lemma 4.6 of Chapter 11 explain why we can write $s(x) = \frac{p}{q}t(x)$.

$$f(x)\left(\frac{p}{q}t(x)\right) = g(x).$$

Multiplying through by q, we obtain an equation in $\mathbb{Z}[x]$,

$$f(x)p\,t(x) = q\,g(x).$$

Now consider the content of each side of this equation. On the left, f and t are both primitive, and so the content is p. On the right, the content is qc_g, where c_g is the content of g. The two sides must be equal, and so $p = qc_g$, but this implies that $\frac{p}{q} = c_g$ is an integer. Therefore $s(x) = c_g t(x)$, and so $s \in \mathbb{Z}[x]$ and hence f divides g in $\mathbb{Z}[x]$. ■

Combining Lemma 4.8 with Proposition 4.9, we conclude the following.

Theorem 4.11

The ring $\mathbb{Z}[x]$ is a unique factorisation domain.

Exercise 4.3

(a) Prove that if f is a prime polynomial in $\mathbb{Z}[x]$ and $\deg(f) \geq 1$, then f is prime in $\mathbb{Q}[x]$.

(b) What happens if $f \in \mathbb{Z}[x]$ is prime, but $\deg(f) = 0$: is f still prime in $\mathbb{Q}[x]$?

4.3 Sums of two squares, revisited

In Subsection 3.3, we established that a prime number p can be expressed as the sum of two squares if, and only if, $p = 2$, or p is of the form $4k + 1$ for some $k \in \mathbb{N}$. We also found, in Worked Exercise 3.12 and Exercise 3.4, two examples of integers that can be expressed as a sum of squares in more than one way. This is not true of primes. No prime of the form $4k + 3$ can be expressed as a sum of two squares, whilst each prime of the form $4k + 1$ can be expressed as a sum of two squares *in a unique way*. The uniqueness assumes, as we have been doing all along, that by squares we mean squares of non-negative integers, and also that the sum of squares $x^2 + y^2$ is regarded as the same expression as $y^2 + x^2$.

In Theorem 3.13 we used the ring of Gaussian integers $\mathbb{Z}[i]$, and the fact that it is a Euclidean domain, to prove the existence of $x, y \in \mathbb{Z}$ such that $p = x^2 + y^2$ whenever p is of the form $4k + 1$. By Theorem 4.5 we now know that $\mathbb{Z}[i]$ is in fact a UFD, and this will enable us to prove the uniqueness of this expression.

Theorem 4.12 *Uniqueness of representation*

The expression of a prime p of the form $4k + 1$ as a sum of two squares is unique except for the order of the two summands.

Proof Suppose that $p = a^2 + b^2 = c^2 + d^2$, where $a, b, c, d \in \mathbb{Z}$. We must show that $a = \pm c$, $b = \pm d$ or $a = \pm d$, $b = \pm c$ so that the expression for p as a sum of two squares is unique (up to associates).

In $\mathbb{Z}[i]$, we have $p = (a + bi)(a - bi) = (c + di)(c - di)$. We claim that $a + bi$ and $a - bi$ are irreducible in $\mathbb{Z}[i]$. With the usual norm, we have $N(a + bi) = |a + bi|^2 = a^2 + b^2 = p$. Since p is prime in $\mathbb{Z}$, the only divisors of p in $\mathbb{Z}$ are 1 and p. Using the fact that N is multiplicative, for any factorisation of $a + bi = w \cdot z$ with $w, z \in \mathbb{Z}[i]$, either $N(w) = 1$ or $N(z) = 1$, and so one of w or z is a unit. Therefore $a + bi$ is irreducible. By using similar reasoning, it can also be shown that $a - bi$, $c + di$ and $c - di$ are irreducible in $\mathbb{Z}[i]$.

Now, since $\mathbb{Z}[i]$ is a UFD, the factorisation of $p = (a + bi)(a - bi)$ into irreducibles is unique up to associates, and thus either $(a + bi) = u(c + di)$ or $(a + bi) = v(c - di)$ where $u, v \in \{1, -1, i, -i\}$ are units. In either case, we compare real and imaginary coefficients to conclude that either $a = \pm c, b = \pm d$ or $a = \pm d, b = \pm c$. ■

We now know exactly which primes can be expressed as a sum of two squares, namely 2 and any prime of the form $4k + 1$, and we also know that any product of numbers expressible as a sum of two squares is itself so expressible. Hence any number that has no prime divisor of the form $4k + 3$ is certainly expressible as a sum of two squares. But are such numbers the only ones? A little experimentation reveals that they are not. For example, $18 = 3^2 + 3^2$ is a sum of two squares that does not adhere to the above prescription because it has a prime divisor of the form $4k + 3$, namely 3 itself. The essential point with 18 is that it is immaterial that 3 cannot be written as a sum of two squares, because 3^2 divides 18 and

$$18 = 2 \times 3^2 = (1^2 + 1^2)3^2 = 3^2 + 3^2.$$

In general, the identity

$$c^2(a^2 + b^2) = (ca)^2 + (cb)^2$$

shows that we can multiply any sum of squares by *any* square and retain a sum of squares. We can build on this observation in the following way. Suppose, for example, that we want to find one way of expressing

$2^7 \times 3^4 \times 5 \times 13^3$ as a sum of squares. By first isolating as large a square term as possible, we can write

$$2^7 \times 3^4 \times 5 \times 13^3 = (2^3 \times 3^2 \times 13)^2 \times (2 \times 5 \times 13)$$

Remember that a square-free integer is one that is not divisible by the square of any prime. Every positive integer is square, square-free or can be expressed as a square multiplied by a square-free integer.

and the task will be completed when we express the square-free part, $2 \times 5 \times 13$, as a sum of two squares.

$$\begin{aligned}(2^3 \times 3^2 \times 13)^2 \times 2 \times 5 \times 13 &= (2^3 \times 3^2 \times 13)^2 \times (1^2+1^2)(1^2+2^2)(2^2+3^2)\\ &= (2^3 \times 3^2 \times 13)^2 \times (1^2+1^2)(8^2+1^2)\\ &= (2^3 \times 3^2 \times 13)^2 \times (9^2+7^2)\\ &= (2^3 \times 3^4 \times 13)^2 + (2^3 \times 3^2 \times 7 \times 13)^2\end{aligned}$$

We make use of this idea in our proof of the following result, which gives a complete classification of which integers are sums of two squares.

Proposition 4.13 *Sums of two squares*

A positive integer n can be expressed as a sum of two squares if, and only if, each of its prime divisors of the form $4k+3$, if any, occurs to an even power.

Proof By pulling out the largest square divisor of n we can write $n = m^2 r$, where r is square-free. The theorem asserts that n is expressible as a sum of two squares if, and only if, r is not divisible by any prime of the form $4k+3$.

We allow the possibility $r = 1$.

To prove the 'if, and only if' assertion we have to establish the implications both ways; so we break the proof into two parts.

(a) Suppose that r has no prime divisor of the form $4k+3$. If $r = 1$ then $n = m^2 + 0^2$ and there is nothing to prove. If $r > 1$ then r is a product of one or more primes, each of which is either 2 or of the form $4k+1$. We have seen that such a product r can be expressed as a sum of two squares, and so $n = m^2(a^2+b^2) = (ma)^2 + (mb)^2$.

(b) Suppose that n can be expressed as a sum of two squares, say

$$n = m^2 r = a^2 + b^2.$$

First, any common divisor of a and b may be cancelled as follows. If $\mathrm{hcf}(a,b) = d$ then we can write $a = a_1 d$, $b = b_1 d$, where $\mathrm{hcf}(a_1, b_1) = 1$, and

$$m^2 r = d^2(a_1^2 + b_1^2).$$

Any prime that divides d occurs with exponent 2, or more, on the right-hand side of this equation and so must divide m.

As r is square-free, d divides m. Writing $m_1 = \frac{m}{d}$ gives

$$m_1^2 r = a_1^2 + b_1^2.$$

Our task is to show that r does not have a prime divisor of the form $4k+3$. Aiming for a contradiction, suppose that the prime $p = 4k+3$ divides r. Then

$$a_1^2 + b_1^2 \equiv 0 \,(\mathrm{mod}\ p); \quad \text{that is, } a_1^2 \equiv -b_1^2 \,(\mathrm{mod}\ p).$$

Now if p divides a_1 we would have p divides b_1, thereby contradicting $\text{hcf}(a_1, b_1) = 1$. So $\text{hcf}(a_1, p) = \text{hcf}(b_1, p) = 1$ and FLT can be applied, giving

$$a_1^{p-1} \equiv b_1^{p-1} \equiv 1 \,(\text{mod } p).$$

Putting $p = 4k + 3$,

$$\begin{aligned} 1 \equiv a_1^{p-1} &= a_1^{4k+2} = (a_1^2)^{2k+1} \equiv (-b_1^2)^{2k+1} \\ &\equiv (-1)^{2k+1}(b_1^2)^{2k+1} \equiv (-1)b_1^{p-1} \equiv -1 \,(\text{mod } p). \end{aligned}$$

This cannot possibly be true for an odd prime p and so we have a contradiction, and the result follows. ■

Exercise 4.4

(a) Which integers in the range 2000 to 2010 inclusive can be expressed as a sum of two squares? For any that can be so expressed, find one such representation.

(b) Use Proposition 4.13 to show that no positive integer that is congruent to either 3 or 6 modulo 9 can be expressed as a sum of two squares. For each of the other seven congruence classes modulo 9, find the smallest positive integer that can be represented as a sum of two squares and the smallest that cannot.

4.4 More Diophantine equations

We finish with two more applications of UFDs to solving Diophantine equations. First, using $\mathbb{Z}[\sqrt{-2}]$, we have the following result.

Proposition 4.14

The equation $a^2 + 2 = b^3$ has exactly one positive integer solution, namely $a = 5, b = 3$.

Proof There are two parts to this proof. First, we show

$$\text{hcf}(a + \sqrt{-2}, a - \sqrt{-2}) = 1;$$

so both $a - \sqrt{-2}$ and $a + \sqrt{-2}$ are cubes in $\mathbb{Z}[\sqrt{-2}]$. Next, we use the fact that

$$a - \sqrt{-2} = (c + d\sqrt{-2})^3$$

to conclude that $a = 5$ and $b = 3$.

We will work in the Euclidean domain $\mathbb{Z}[\sqrt{-2}]$. Here, $a^2 + 2 = (a + \sqrt{-2})(a - \sqrt{-2})$. We first claim that $\text{hcf}(a + \sqrt{-2}, a - \sqrt{-2}) = 1$.

Heading for a contradiction, suppose that p is an irreducible (and hence prime) element of $\mathbb{Z}[\sqrt{-2}]$ dividing $a+\sqrt{-2}$ and $a-\sqrt{-2}$. Then p divides $(a+\sqrt{-2})-(a-\sqrt{-2}) = 2\sqrt{-2} = (\sqrt{-2})^3$. By considering norms, $\sqrt{-2}$ is irreducible in $\mathbb{Z}[\sqrt{-2}]$, and since p is also irreducible, unique factorisation means that $p = \sqrt{-2}$ (or $p = -\sqrt{-2}$).

Now, $p = \sqrt{-2}$ divides $a+\sqrt{-2}$, and so $N(p) = 2$ divides $N(a+\sqrt{-2}) = a^2+2$. In other words, a^2+2 is even, and therefore a must be even. This in turn implies that $a^2+2 = 4k+2$ for some integer k. Next, from $a^2+2 = b^3$, we find that b^3 must be even. The only way for this to happen is if b is even, and then b^3 is of the form 8ℓ for some integer ℓ. Thus $4k+2 = 8\ell$. Dividing by 2 we find $2k+1 = 4\ell$, but the left-hand side is odd and the right-hand side is even, which is impossible. Thus there can be no irreducible element p dividing both $a+\sqrt{-2}$ and $a-\sqrt{-2}$, and hence $\text{hcf}(a+\sqrt{-2}, a-\sqrt{-2}) = 1$.

Finally, given $(a+\sqrt{-2})(a-\sqrt{-2}) = b^3$, and since $a+\sqrt{-2}$ and $a-\sqrt{-2}$ have a highest common factor of 1 in $\mathbb{Z}[\sqrt{-2}]$, they must both themselves be cubes. Then we have $a-\sqrt{-2} = (c+d\sqrt{-2})^3$ for some $c, d \in \mathbb{Z}$, which gives us

$$a-\sqrt{-2} = (c^3-6cd^2)+(3c^2d-2d^3)\sqrt{-2}.$$

Comparing coefficients, we have

$$\begin{aligned} a &= c^3-6cd^2 \\ -1 &= 3c^2d-2d^3 = d(3c^2-2d^2). \end{aligned}$$

The second equation implies $d = \pm 1$, which leaves us with $3c^2-2 = \mp 1$. If $d=-1$, then $3c^2 = 1$ has no solution in the integers, and so we must have $d=1$. Therefore $3c^2 = 3$, and so $c = \pm 1$.

Now, using the equation $a = c^3-6cd^2$, we have $a = (\pm 1)^3 \mp 6\cdot 1\cdot 1 = \mp 5$. Taking the positive solution, we get $a=5$, and then $b^3 = 27$ implies $b=3$, as required. ■

We finish with an exercise for you to try, which will lead you through the steps to prove that $a^2+1 = b^3$ has no non-trivial solutions in $\mathbb{Z}$.

Exercise 4.5

Let $a, b \in \mathbb{Z}$.

(a) Prove that if $a^2+1 = b^3$, then b must be odd, and hence a is even.

(b) Show that $1+i$, $1-i$ are irreducible elements in $\mathbb{Z}[i]$.

(c) Using parts (a) and (b), show that $\text{hcf}(a+i, a-i) = 1$ in $\mathbb{Z}[i]$.

Hint: use the fact that $2 = (1+i)(1-i)$.

(d) Hence conclude that the only solution to $a^2+1 = b^3$ in $\mathbb{Z}$ is $a=0$ and $b=1$.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should:

(a) Understand the term Pythagorean triple, recognise when a triple is primitive, and understand the statement and proof classifying the primitive Pythagorean triples.

(b) Understand the method of infinite descent, and be able to apply it to prove that certain Diophantine equations have no solutions.

(c) Know the meaning of the terms associate, factor, irreducible and prime as applied to elements of an arbitrary ring.

(d) Know what is meant by an integral domain, and understand why in an integral domain every prime element is irreducible, but not every irreducible element is prime.

(e) Recall the meaning of norm, multiplicative norm and Euclidean norm, apply these concepts to prove the irreducibility of ring elements, and use this to establish that in certain rings factorisation is not unique.

(f) Know the meaning of Euclidean domain, recall the approach to show that $\mathbb{Z}[i]$ is a Euclidean domain by finding the nearest point to a ring element, and be able to apply this approach to other rings.

(g) Know the definition of highest common factor of two elements of a ring, recall the approach to prove that highest common factors exist in Euclidean domains, and thus be able to show that certain rings are not Euclidean domains.

(h) Understand the steps required to prove that irreducible and prime elements coincide in a Euclidean domain.

(i) Understand the role Euclidean domains play in classifying when a prime number can be expressed as a sum of two squares, and use the method of exhaustion to express integers as a sum of two squares.

(j) Recall the definition of unique factorisation domain, the proof that every irreducible element is prime in a UFD, and that every Euclidean domain is a UFD.

(k) Be able to show that certain rings are not Euclidean domains by proving that they are not unique factorisation domains.

(l) Understand the key steps in the proof that $\mathbb{Z}[x]$ is a unique factorisation domain.

(m) Know why the fact that $\mathbb{Z}[i]$ is a UFD implies that primes of the form $4k+1$ are uniquely expressible as a sum of two squares except for the order of the two summands, and apply methods to show when a positive integer can or cannot be expressed as a sum of two squares.

(n) Be able to apply the theory of UFDs to compute all solutions (or show that none exist) to certain Diophantine equations.

Solutions to exercises

Solution to Exercise 1.1

(a) $(10, 8, 6)$ is not a Pythagorean triple; the hypotenuse must be the largest of the three values and, by convention, must be the last member of the triple.

(b) $(12, 9, 15)$ is a Pythagorean triple, but it is not primitive since it is the $(4, 3, 5)$ triple scaled by a multiple of 3.

(c) $(6, 7, 8)$ is not a Pythagorean triple: $6^2 + 7^2 \neq 8^2$.

(d) $(5, 12, 13)$ is a primitive Pythagorean triple: $5^2 + 12^2 = 169 = 13^2$ and the three integers are pairwise coprime.

(e) $(24, 33, 41)$ is not a Pythagorean triple since 24 and 33 have a common factor, namely 3, that does not divide the third number.

Solution to Exercise 1.2

(a) We seek coprime natural numbers m and n, of opposite parity and with $m > n$, such that $2mn = 72$. There are two such pairs:

$m = 36,\ n = 1,$ giving the triple $(72, 1295, 1297)$;

$m = 9,\ n = 4,$ giving the triple $(72, 65, 97)$.

The pairs $m = 18$, $n = 2$ and $m = 12$, $n = 3$ are excluded because they are not coprime.

(b) One possible strategy is to look at each divisor of 30 in turn and to determine all primitive triples containing that divisor. But notice that, since 30 is not a multiple of 4, no even divisor of 30 (including 30 itself) can occur in a primitive Pythagorean triple. So we need look only at the odd divisors of 30. Any primitive triple containing one of 3, 5 or 15 may be scaled appropriately to give a Pythagorean triple containing 30.

Note that the even member in a primitive triple is divisible by 4, since mn is even and $x = 2mn$.

We therefore seek all pairs of coprime natural numbers m and n of opposite parity and with $m > n$ such that either $m^2 + n^2$ or $m^2 - n^2$ is equal to one of 3, 5 or 15. It turns out that there are just five such pairs, as presented below.

Divisor	m	n	Primitive triple	Scaling	Triple with 30
15	4	1	$(8, 15, 17)$	2	$(16, 30, 34)$
15	8	7	$(112, 15, 113)$	2	$(224, 30, 226)$
5	3	2	$(12, 5, 13)$	6	$(72, 30, 78)$
5	2	1	$(4, 3, 5)$	6	$(24, 18, 30)$
3	2	1	$(4, 3, 5)$	10	$(40, 30, 50)$

As $m^2 - n^2 = (m - n)(m + n)$ we need only consider m and n for which $m + n \leq 15$.

(c) Suppose that $st = u^2$, where s and t are coprime. Let s and t have prime decompositions $s = p_1^{k_1} p_2^{k_2} \cdots p_i^{k_i}$, $t = p_{i+1}^{k_{i+1}} p_{i+2}^{k_{i+2}} \cdots p_j^{k_j}$. As $\mathrm{hcf}(s,t) = 1$, the listed primes are all distinct and so st has prime decomposition

$$st = p_1^{k_1} \cdots p_i^{k_i} p_{i+1}^{k_{i+1}} \cdots p_j^{k_j}.$$

The primes are not necessarily listed in ascending order here.

As st is a square, each exponent in this expression is even. Hence each exponent in s, and each exponent in t, is even and so these two are squares as well.

Solution to Exercise 1.3

Suppose that $x_1^4 + 2y_1^4 = 4z_1^4$, where x_1, y_1 and z_1 are positive integers. As the prime 2 divides two of the terms it must divide the third, namely x_1^4. So 2 divides x_1 and we can write $x_1 = 2x_2$ for some positive integer x_2. Substituting for x_1 gives

$$16x_2^4 + 2y_1^4 = 4z_1^4; \quad \text{that is, } 8x_2^4 + y_1^4 = 2z_1^4.$$

Repeating the argument, we can write $y_1 = 2y_2$ for some positive integer y_2 and obtain

$$8x_2^4 + 16y_2^4 = 2z_1^4; \quad \text{that is, } 4x_2^4 + 8y_2^4 = z_1^4.$$

We can now write $z_1 = 2z_2$ for some positive integer z_2 to obtain

$$4x_2^4 + 8y_2^4 = 16z_2^4; \quad \text{that is, } x_2^4 + 2y_2^4 = 4z_2^4.$$

At this point we have reached a second positive solution, $x = x_2$, $y = y_2$, $z = z_2$ of the original equation with $z_2 < z_1$. The descent step is therefore complete and the required contradiction is established.

Thus $x^4 + 2y^4 = 4z^4$ has no positive solutions.

Solution to Exercise 1.4

(a) As any square is congruent modulo 4 to 0 (if it is even) or to 1 (if it is odd), it follows that $x^2 + y^2 + z^2 \equiv 0 \pmod 4$ can occur only when x, y and z are all even.

(b) Suppose that $x_1^2 + y_1^2 + z_1^2 = 2x_1y_1z_1$. If x_1, y_1 and z_1 are all odd then $x_1^2 + y_1^2 + z_1^2 \equiv 3 \pmod 4$ while the right-hand side, $2x_1y_1z_1$, is even. Hence at least one of x_1, y_1 and z_1 must be even. But then $x_1^2 + y_1^2 + z_1^2 = 2x_1y_1z_1 \equiv 0 \pmod 4$ and, as we have just seen, x_1, y_1 and z_1 must all be even.

Writing $x_1 = 2x_2$, $y_1 = 2y_2$ and $z_1 = 2z_2$ gives

$$(2x_2)^2 + (2y_2)^2 + (2z_2)^2 = 2(2x_2)(2y_2)(2z_2),$$

which simplifies to

$$x_2^2 + y_2^2 + z_2^2 = 4x_2y_2z_2.$$

As the right-hand side is congruent modulo 4 to 0, the first part of the question once again gives that x_2, y_2, and z_2 are even positive integers.

(c) Continuing from part (b), if we write $x_2 = 2x_3$, $y_2 = 2y_3$, $z_2 = 2z_3$ we obtain

$$x_3^2 + y_3^2 + z_3^2 = 8x_3y_3z_3$$

with x_3, y_3 and z_3 being even positive integers. We can continue forever in this way, halving the x, y and z values and yet retaining even positive integers, for from

$$x_n^2 + y_n^2 + z_n^2 = 2^n x_n y_n z_n$$

we note that x_n, y_n and z_n are even integers and

$$\left(\frac{x_n}{2}\right)^2 + \left(\frac{y_n}{2}\right)^2 + \left(\frac{z_n}{2}\right)^2 = 2^{n+1}\left(\frac{x_n}{2}\right)\left(\frac{y_n}{2}\right)\left(\frac{z_n}{2}\right).$$

The same argument now confirms that $\frac{x_n}{2}$, $\frac{y_n}{2}$ and $\frac{z_n}{2}$ are still even positive integers. As the sequence $z_1, z_2, z_3, \ldots$ of positive integers cannot decrease indefinitely, the method of infinite descent shows that the supposed positive solution of the original equation cannot exist.

Solution to Exercise 1.5

From $x^2 + y^2 = z^2$ and $xy = 2n^2$ we have

$$(x+y)^2 = x^2 + y^2 + 2xy = z^2 + (2n)^2$$

and

$$(x-y)^2 = x^2 + y^2 - 2xy = z^2 - (2n)^2.$$

Multiplying these two equations together gives

$$(x^2 - y^2)^2 = z^4 - (2n)^4.$$

Note that $x \neq y$ in any Pythagorean triple and that $x^2 - y^2$ can be replaced by $y^2 - x^2$ if the former is negative.

By Corollary 1.10 no positive solution can exist. We conclude that no Pythagorean triangle can have an area that is a square.

Solution to Exercise 2.1

(a) If a, b satisfy $a^2 - db^2 = 1$, then in $\mathbb{Z}[\sqrt{d}]$ we have $(a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - db^2 = 1$, and so $a + b\sqrt{d}$ is a unit.

(b) Since a, b are both greater than 0, the unit $u = a + b\sqrt{d}$ of $\mathbb{Z}[\sqrt{d}]$ is a real number that is strictly greater than 1. By Lemma 2.2, u^k is a unit of $\mathbb{Z}[\sqrt{d}]$ for all $k \geq 1$, and since $u > 1$ these are all distinct. Hence $\mathbb{Z}[\sqrt{d}]$ has infinitely many units when $d > 1$ is square-free.

Solution to Exercise 2.2

(a) (i) Using the multiplication table for $\mathbb{Z}_6$ given in Example 2.7, we find that the units are 1 and 5, since $5 \cdot 5 \equiv 1 \pmod 6$.

The non-zero non-units are therefore 2, 3 and 4. None is irreducible since $2 = 2 \cdot 4$, $3 = 3 \cdot 3$ and $4 = 2 \cdot 2$, and so there are no irreducible elements in $\mathbb{Z}_6$. However, all of 2, 3 and 4 are prime. Example 2.7 shows that 2 is prime. Next, 3 divides 0 and 3, but every product $a \cdot b \equiv 0$ or $3 \pmod 6$ requires that one of a or b is 0 or 3. Finally, we note that 4 divides 0, 2 and 4, but whenever $a \cdot b \equiv 0, 2$ or $4 \pmod 6$ we find that at least one of a or b is 0, 2 or 4.

(ii) The multiplication table for $\mathbb{Z}_8 \setminus \{0\}$ is:

	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	2	4	6	0	2	4	6
3	3	6	1	4	7	2	5
4	4	0	4	0	4	0	4
5	5	2	7	4	1	6	3
6	6	4	2	0	6	4	2
7	7	6	5	4	3	2	1

The units are $1, 3, 5$ and 7, since $3 \cdot 3 \equiv 1$, $5 \cdot 5 \equiv 1$ and $7 \cdot 7 \equiv 1 \pmod 8$.

The non-zero non-units are therefore 2, 4 and 6. The irreducible elements are 2 and 6 since whenever $a \cdot b \equiv 2$ or $6 \pmod 8$, one of a or b is a unit. 4 is not irreducible since $4 = 2 \cdot 2$.

The primes in $\mathbb{Z}_8$ are also 2 and 6. First, 4 is not prime since $4 = 2 \cdot 2$, but reading along the 4th row we see that 4 does not divide 2. To prove that 2 and 6 are prime, observe that both 2 and 6 divide the elements 0, 2, 4 and 6. Whenever $a \cdot b \equiv 0$, 2, 4 or $6 \pmod 8$ then one of a or b is equal to one of 0, 2, 4 or 6.

(b) No, since $(4 + 3i) = (1 + 2i)(2 - i)$. Do not worry if you did not find this factorisation. The point of this question was to make you aware that deciding irreducibility is not very easy.

Notice that

$$\begin{aligned} |4 + 3i|^2 &= 4^2 + 3^2 = 25 \\ &= 5 \cdot 5 = (1 + 4) \cdot (4 + 1) \\ &= |1 + 2i|^2 \cdot |2 - i|^2, \end{aligned}$$

which suggests how you might go about finding such obscure factorisations.

Solution to Exercise 2.3

(a) Since $a \mid c$, there exists $b \in R$ such that $c = ba$. Now, $c + na = ba + na = (b + n)a$ proves that a divides $c + na$.

(b) Since $a \mid b$, there exists non-zero $c \in R$ such that $b = ac$. Similarly, $b \mid a$ implies there exists non-zero $d \in R$ such that $a = bd$. Now $a = bd = acd$ and Lemma 2.11 now shows that $cd = 1$. So c and d are units, and hence a and b are associates.

(c) Since $a \mid b$ and $b \mid c$, there exist non-zero $p, q \in R$ such that $b = ap$ and $c = bq$. Now $c = bq = (ap)q = a(pq)$ and therefore a divides c.

(d) Since $a \mid b$ and $a \mid c$, there exist non-zero $p, q \in R$ such that $b = ap$ and $c = aq$. Now $mb + nc = map + naq = amp + anq = a(mp + nq)$ and so a divides $mb + nc$.

(e) This follows since $0 = a \cdot 0$ for all $a \in R$.

(f) This follows from $n = u(u^{-1}n)$.

Solution to Exercise 2.4

Suppose $S \subseteq R$ is not an integral domain. Then there exist non-zero $a, b \in S$ such that $a \cdot b = 0_S$, where 0_S denotes the additive identity of the ring S.

Since S is a subring of R, we know that $0_S = 0_R$. Therefore $a \cdot b = 0_R$ in R, and since $a, b \neq 0_R$ this shows that R cannot be an integral domain if S is not an integral domain.

Solution to Exercise 2.5

(a) Since R is a subring of the integral domain $\mathbb{C}$, it follows by Exercise 2.4 that R is an integral domain.

Alternatively, notice that N is the complex modulus squared, which we know to be multiplicative.

(b) Let $r = a + b\sqrt{-5}, s = c + d\sqrt{-5} \in R \setminus \{0\}$. Then

$$\begin{aligned} N(rs) &= N((ac - 5bd) + (bc + ad)\sqrt{-5}) \\ &= (ac - 5bd)^2 + 5(bc + ad)^2 \\ &= a^2c^2 - 10abcd + 25b^2d^2 + 5b^2c^2 + 10abcd + 5a^2d^2 \\ &= a^2c^2 + 25b^2d^2 + 5b^2c^2 + 5a^2d^2 \\ &= (a^2 + 5b^2)(c^2 + 5d^2) = N(r)N(s). \end{aligned}$$

(c) The result follows immediately since if $b \neq 0$ then $b^2 \geq 1$ and so $a^2 + 5b^2 \geq 5$. If $N(r) < 5$ then $b = 0$, and so $r \in \mathbb{Z} \setminus \{0\}$. In fact, the only possibilities are $r = \pm 1, \pm 2$.

(d) Suppose $2 = rs$ for some $r, s \in R$. By part (b), $N(r)$ and $N(s)$ both divide $N(2) = 4$. Therefore, we have $N(r), N(s) < 5$, and so by part (c) r and s must be integers. However, 2 does not have any proper integer divisors, and so in R it is irreducible.

(e) If 2 divided $(1+\sqrt{-5})$ or $(1-\sqrt{-5})$ then, since N is multiplicative, $N(2)$ would divide $N(1+\sqrt{-5})$ or $N(1-\sqrt{-5})$. However, $N(2)=4$ does not divide $N(1+\sqrt{-5})=N(1-\sqrt{-5})=6$, and so 2 divides neither $(1+\sqrt{-5})$ nor $(1-\sqrt{-5})$.

This proves that 2 is not prime, since 2 divides $2\cdot 3=6$ $=(1+\sqrt{-5})(1-\sqrt{-5})$, but it divides neither $(1+\sqrt{-5})$ nor $(1-\sqrt{-5})$.

Solution to Exercise 2.6

(a) If $u=a+b\sqrt{-5}$ is a unit in R, then there is $u^{-1}\in R$ with $uu^{-1}=1$. This implies that $N(u)N(u^{-1})=N(1)=1$. Since $N(u),N(u^{-1})\geq 1$ we must have $N(u)=1$. Now $N(u)=N(a+b\sqrt{-5})=a^2+5b^2$, and the only solutions to $a^2+5b^2=1$ are $a=\pm 1, b=0$. Therefore, the only elements in R with $N(u)=1$ are $1,-1$.

Note that since N is defined only for non-zero elements of R, we have $N(a+b\sqrt{-5})>0$.

Finally, $1\cdot 1=1$ and $(-1)\cdot(-1)=1$ proves that both 1 and -1 are indeed units.

(b) First, we compute the norm: $N(3)=3^2+0=9$. If 3 is reducible, then $3=rs$ where $r,s\in R$ are not units. Since N is multiplicative, this means that $N(r)N(s)=N(rs)=N(3)=9$. Next, r and s are not units, and so $N(r)>1$ and $N(s)>1$ (using part (a)), and therefore $N(r)=N(s)=3$.

From Exercise 2.5(c), the only non-zero elements of R for which $N(r)<5$ are integers, with $r=\pm 1,\pm 2$. Now, $N(\pm 1)=1$ and $N(\pm 2)=4$, and so there are no elements $r\in R$ with $N(r)=3$. Therefore we cannot have $3=rs$, unless one of r or s is a unit.

(c) Suppose that $r,s\in R$ are not units, with $1+\sqrt{-5}=rs$. Then $N(rs)=N(1+\sqrt{-5})=6=N(r)N(s)$. Since $N(r)>1$ and $N(s)>1$, the norms of r and s must be 2 and 3, in some order. However, in part (b) we proved that there are no elements of norm 3, and so this is impossible. Thus $1+\sqrt{-5}$ is irreducible.

The same argument shows that $1-\sqrt{-5}$ is irreducible.

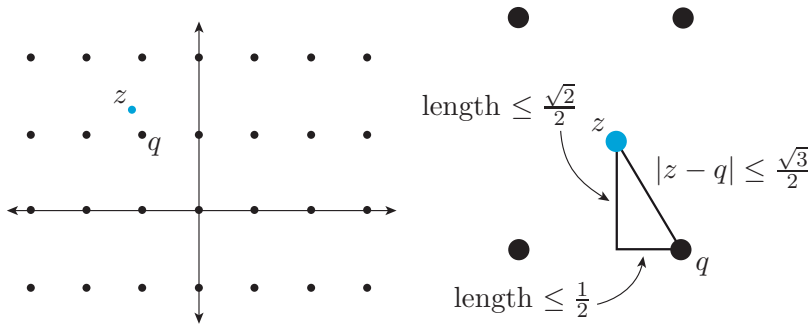
(d) $6=(1+\sqrt{-5})(1-\sqrt{-5})=2\cdot 3$, where all four factors are irreducible, and none is an associate of another since the only units are 1 and -1.

Solution to Exercise 3.1

(a) We follow the approach laid out for $\mathbb{Z}[i]$ in Worked Exercise 3.4. First, $\mathbb{Z}[\sqrt{-2}]$ is an integral domain since it is a subring of $\mathbb{C}$.

We now show that N is a Euclidean norm. For non-zero $a,b\in\mathbb{Z}[\sqrt{-2}]$, we have $N(ab)=N(a)N(b)\geq N(a)$ since $N(b)\geq 1$ unless $b=0$.

Taking $a,b\in\mathbb{Z}[\sqrt{-2}]$ and $b\neq 0$, to find q,r such that $a=bq+r$ we set q to be a point in $\mathbb{Z}[\sqrt{-2}]$ that is nearest to the complex number $z=\frac{a}{b}$.



The distance from z to q in this diagram is $|z-q|$. Since q is the nearest point of $\mathbb{Z}[\sqrt{-2}]$ to z, $|z-q|$ cannot be more than the distance from q to the centre of the rectangle bounded by points of $\mathbb{Z}[\sqrt{-2}]$ containing z. This rectangle has dimensions $1 \times \sqrt{2}$, and so the distance from a corner to the centre is $\frac{\sqrt{3}}{2}$. Therefore

$$|z-q| \leq \tfrac{\sqrt{3}}{2} < 1.$$

Now, set $r = a - bq$. Then $r \in \mathbb{Z}[\sqrt{-2}]$, and

$$N(r) = |r|^2 = |a-bq|^2 = |b|^2|z-q|^2 < |b|^2 = N(b),$$

which proves that N is a Euclidean norm on $\mathbb{Z}[\sqrt{-2}]$.

Note that 'failing to prove $\mathbb{Z}[\sqrt{-5}]$ is Euclidean' does not prove anything: it merely indicates why the earlier approach does not work.

(b) If we attempt to prove that $\mathbb{Z}[\sqrt{-5}]$ is a Euclidean domain by modifying the argument we used for part (a) of this exercise, we find that the points $0, \sqrt{-5}, 1, 1+\sqrt{-5}$ of $\mathbb{Z}[\sqrt{-5}]$ form the corners of a rectangle in the complex plane whose sides have lengths 1 and $\sqrt{5} \approx 2.236$. This implies that the centre of the rectangle is at a distance $\frac{\sqrt{6}}{2} \approx 1.223 > 1$ from any of the corners.

Thus, for $a, b \in \mathbb{Z}[\sqrt{-5}]$, with $b \neq 0$, it is possible that if $q \in \mathbb{Z}[\sqrt{-5}]$ is the nearest point to the complex number $z = \frac{a}{b}$, then we have $|z-q| > 1$. If this occurs, then setting $r = a - bq$ as before does not work, since $N(r) = N(a-bq) = |b|^2|z-q|^2 > N(b)$.

(c) Suppose there exist $q, r \in \mathbb{Z}[\sqrt{-5}]$ such that $1+\sqrt{-5} = 2q + r$ with either $r = 0$ or $N(r) < N(2)$. Because $N(2) = 2^2 = 4$ and $N(1+\sqrt{-5}) = 1+5 = 6$, we know that 2 does not divide $1+\sqrt{-5}$ and so $r \neq 0$. We therefore must have $N(r) < N(2) = 4$. From Exercise 2.5, the condition $N(r) < 4$ implies that $r = 1$ or -1.

Write $q = a + b\sqrt{-5}$ with $a, b \in \mathbb{Z}$. If $r = 1$ then $1+\sqrt{-5} = 2q+1$, and subtracting 1 from either side gives us

$$\sqrt{-5} = 2a + 2b\sqrt{-5}$$

from which we conclude $a = 0$ and $b = \frac{1}{2}$. However, b must be an integer, and so this is impossible.

Similarly, if $r = -1$, then $1+\sqrt{-5} = 2q - 1$, which tells us

$$2 + \sqrt{-5} = 2a + 2b\sqrt{-5}$$

and so $a = 1$, but again $b = \frac{1}{2}$, which is impossible. Therefore no such r can exist with $N(r) < N(2)$, and so N is not a Euclidean norm.

Solution to Exercise 3.2

(a) For non-zero $a+bi \in \mathbb{Z}[i]$, we will use the Euclidean norm $N(a+bi) = |a+bi|^2 = a^2+b^2$. From the fact that $N(1) = 1$, it follows that if $a+bi$ is a unit in $\mathbb{Z}[i]$ then it must satisfy $a^2+b^2 = 1$.

If $|a| > 1$ or $|b| > 1$, then $a^2+b^2 \geq 4$ and so there is no solution to $a^2+b^2 = 1$. We cannot have $a = b = 0$, or $|a| = |b| = 1$, and this leaves four cases: $a = \pm 1$ and $b = 0$, or $a = 0$ and $b = \pm 1$. Hence, the units in $\mathbb{Z}[i]$ are $\pm 1, \pm i$.

(b) Since $1 \in \mathbb{Z}[\sqrt{-2}]$ is a unit and $N(1) = 1^2 + 2 \cdot 0^2 = 1$, any unit $a+b\sqrt{-2}$ must satisfy $a^2+2b^2 = 1$. If $b \neq 0$ then $a^2+2b^2 > 1$, and so we must have $b = 0$. This leaves $a^2 = 1$, and so the only units of $\mathbb{Z}[\sqrt{-2}]$ are 1 and -1.

Solution to Exercise 3.3

(a) The only divisors of $2 \in \mathbb{Z}[x]$ are 1 and 2, and the divisors of $x \in \mathbb{Z}[x]$ are 1 and x. Therefore, 1 is a common factor, and is clearly maximal.

(b) If $\mathbb{Z}[x]$ is a Euclidean domain, then since $\text{hcf}(2, x) = 1$ there should exist polynomials $s, t \in \mathbb{Z}[x]$ such that

$$2s(x) + xt(x) = 1.$$

Notice that the constant term on the left-hand side of this equation must equal 1. However, since $x \cdot t(x)$ has constant term zero, the constant term of $2s(x)$ must equal 1, but this is impossible since all coefficients of $2s(x)$ must be divisible by 2. This tells us that $\mathbb{Z}[x]$ is not a Euclidean domain.

(c) By inspection, we may take $s(x) = \frac{1}{2}$ and $t(x) = 0$.

Solution to Exercise 3.4

$$\begin{aligned} 325 &= 5^2 \times 13 = (2^2+1^2)(2^2+1^2)(3^2+2^2) \\ &= (2^2+1^2)(8^2+1^2) = 17^2+6^2 \\ &= (1^2+2^2)(8^2+1^2) = 10^2+15^2 \\ &= (2^2+1^2)(2^2+3^2)(2^2+1^2) = (7^2+4^2)(2^2+1^2) = 18^2+1^2 \end{aligned}$$

Solution to Exercise 3.5

As $313 \equiv 1 \pmod 4$ we know that 313 can be expressed as a sum of two squares. Searching the sequence

$$313 - 1^2, \quad 313 - 2^2, \quad 313 - 3^2, \quad \ldots$$

for the first square, we find $313 - 12^2 = 13^2$. Therefore

$$5321 = 17 \times 313 = (1^2+4^2)(12^2+13^2) = 64^2+35^2.$$

In fact there is just one alternative solution, which can be reached by writing 313 as 13^2+12^2:

$$5321 = (1^2+4^2)(13^2+12^2) = 61^2+40^2.$$

The expression for a prime is unique, but it may not be so for a composite integer.

Solution to Exercise 4.1

For non-zero $a, b \in R$, since R is a UFD we can write a and b uniquely as products of irreducible elements in R. We fix irreducibles $p_1, p_2, \ldots, p_n$ and a unit $u \in R$ so that we can write

$$a = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}, \qquad b = u p_1^{b_1} p_2^{b_2} \cdots p_n^{b_n}$$

for some exponents $a_i, b_i \geq 0$ (note $a_i = 0$ or $b_i = 0$ is explicitly allowed). Note that the unit u simply recognises the fact that b could be an associate of the product $p_1^{b_1} p_2^{b_2} \cdots p_n^{b_n}$.

We claim $\ell = p_1^{\max(a_1,b_1)} p_2^{\max(a_2,b_2)} \cdots p_n^{\max(a_n,b_n)}$ is a least common multiple for a and b. Certainly a divides ℓ, since each irreducible power $p_i^{a_i}$ divides $p_i^{\max(a_i,b_i)}$, and similarly b divides ℓ.

Now suppose $h \in R$ is divisible by a and b. Then since the factorisation of h is unique, $p_i^{\max(a_i,b_i)}$ divides h for each $i = 1, 2, \ldots, n$, and so ℓ divides h, as required.

Solution to Exercise 4.2

We prove that 2 is irreducible in $\mathbb{Z}[\sqrt{-6}]$, but does not divide $\sqrt{-6}$, so that 2 is irreducible but not prime in $\mathbb{Z}[\sqrt{-6}]$, since $6 = 2 \cdot 3 = -(\sqrt{-6})^2$. Then, by Corollary 4.3 and the comments that follow it, we know that $\mathbb{Z}[\sqrt{-6}]$ is not a UFD, and hence not a Euclidean domain by Theorem 4.5.

Alternatively, you could show that the elements 2, 3 and $\sqrt{-6}$ are all irreducible in $\mathbb{Z}[\sqrt{-6}]$, and therefore $\mathbb{Z}[\sqrt{-6}]$ is not a UFD since $6 = 2 \cdot 3 = -(\sqrt{-6})^2$.

To prove that 2 is irreducible and does not divide $\sqrt{-6}$, we use the multiplicative norm $N(a + b\sqrt{-6}) = |a + b\sqrt{-6}|^2 = a^2 + 6b^2$. Since $N(2) = 4$ and N is multiplicative, any element dividing 2 must have norm dividing 4. If $N(a + b\sqrt{-6}) = 1$, then $a^2 + 6b^2 = 1$ implies that $a = \pm 1$ and $b = 0$, and so the only elements with norm 1 are the units $+1$ and -1. Therefore, if the element 2 is reducible then its two factors in $\mathbb{Z}[\sqrt{-6}]$ must both have norm equal to 2. However, there are no elements with norm 2 in $\mathbb{Z}[\sqrt{-6}]$, as there are no solutions to $a^2 + 6b^2 = 2$ in $\mathbb{Z}$: we must have $b = 0$ otherwise $a^2 + 6b^2 \geq 6$, but this leaves $a^2 = 2$, which has no solution in $\mathbb{Z}$. Thus there are no elements in $\mathbb{Z}[\sqrt{-6}]$ with norm 2, and therefore the element 2 is irreducible in $\mathbb{Z}[\sqrt{-6}]$.

Notice that we can't use results from Lemma 3.5 here, because that applies only to Euclidean domains.

Finally, since N is multiplicative and $N(2) = 4$ does not divide $N(\sqrt{-6}) = 6$, we conclude that 2 does not divide $\sqrt{-6}$, and is therefore irreducible but not prime. Hence $\mathbb{Z}[\sqrt{-6}]$ is not a UFD, and hence not a Euclidean domain.

Solution to Exercise 4.3

(a) Since f is prime in $\mathbb{Z}[x]$, it is irreducible in $\mathbb{Z}[x]$ by Theorem 2.12 of this chapter. Since $\deg(f) \geq 1$, f is primitive (otherwise $f = c_f f^*$ is a proper factorisation).

By Gauss's Lemma (Theorem 4.13 of Chapter 11), f is irreducible in $\mathbb{Q}[x]$. Finally, since $\mathbb{Q}[x]$ is a Euclidean domain, every irreducible element is prime, and so f is prime in $\mathbb{Q}[x]$.

(b) No – $\deg(f) = 0$ means that $f \in \mathbb{Z}$, and hence $f \in \mathbb{Q}$. However, $\mathbb{Q}$ is a field, and so f is a unit in $\mathbb{Q}$. All units of $\mathbb{Q}$ are units of $\mathbb{Q}[x]$, and so f cannot be prime in $\mathbb{Q}[x]$.

Solution to Exercise 4.4

(a)

$2000 = 2^4 \times 5^3$	$2000 = 40^2 + 20^2$
$2001 = 3 \times 23 \times 29$	not expressible since $3 \equiv 3 \pmod 4$
$2002 = 2 \times 7 \times 11 \times 13$	not expressible since $7 \equiv 3 \pmod 4$
2003 is a prime	not expressible since $2003 \equiv 3 \pmod 4$
$2004 = 2^2 \times 3 \times 167$	not expressible since $3 \equiv 3 \pmod 4$
$2005 = 5 \times 401$	$2005 = 41^2 + 18^2$
$2006 = 2 \times 17 \times 59$	not expressible since $59 \equiv 3 \pmod 4$
$2007 = 3^2 \times 223$	not expressible since $223 \equiv 3 \pmod 4$
$2008 = 2^3 \times 251$	not expressible since $251 \equiv 3 \pmod 4$
$2009 = 7^2 \times 41$	$2009 = 28^2 + 35^2$
$2010 = 2 \times 3 \times 5 \times 67$	not expressible since $3 \equiv 3 \pmod 4$

(b) If $n \equiv 3 \pmod 9$ then $n = 9k + 3 = 3(3k + 1)$ for some integer k. Now $\text{hcf}(3, 3k + 1) = 1$ and so, in the prime decomposition of n, 3 occurs with exponent 1. Proposition 4.13 confirms that n cannot be expressed as a sum of two squares.

Similarly, if $n = 9k + 6 = 3(3k + 2)$ then n is divisible by 3 but not by 3^2 and so it is not expressible as a sum of two squares.

Taking each of the other congruence classes modulo 9 in turn:

Residue class modulo 9	Smallest sum of two squares	Smallest not sum of two squares
0	$9 = 3^2 + 0^2$	$27 = 3^3$
1	$1 = 1^2 + 0^2$	19
2	$2 = 1^2 + 1^2$	11
4	$4 = 2^2 + 0^2$	$22 = 2 \times 11$
5	$5 = 2^2 + 1^2$	$14 = 2 \times 7$
7	$16 = 4^2 + 0^2$	7
8	$8 = 2^2 + 2^2$	$35 = 5 \times 7$

Each of the primes 3, 7, 11 and 19 is of the form $4k + 3$.

Solution to Exercise 4.5

(a) Suppose, heading for a contradiction, that b is even, and so $b = 2k$, which implies $b^3 = 8k$. From $a^2 + 1 = b^3 = 8k$, we have $a^2 = 8k - 1 = 4(2k - 1) + 3 \equiv 3 \pmod 4$. However, all square numbers are congruent to 0 or 1 modulo 4, and so there is no integer a for which $a^2 = 4(2k - 1) + 3$. Therefore, b must be odd, and so a must be even.

(b) This follows directly by taking norms: $N(1 + i) = |1 + i|^2 = 2$, and since N is multiplicative and 2 is prime in $\mathbb{Z}$, $1 + i$ can have no proper factors. Similarly, $N(1 - i) = 2$ gives the same conclusion.

(c) Suppose p is an irreducible element of $\mathbb{Z}[i]$ that divides $a + i$ and $a - i$. Then p divides $(a + i) - (a - i) = 2i$. Now, i is a unit, and we can write $2 = (1 + i)(1 - i)$. So if p is irreducible and divides $2i$, then by part (b) and the fact that $\mathbb{Z}[i]$ is a UFD we have $p = 1 + i$ or $p = 1 - i$.

If $p = 1 + i$, then since p divides $a + i$, we can find $c, d \in \mathbb{Z}$ such that $(1 + i)(c + di) = a + i$, from which we find $c - d = a$ and $c + d = 1$. Eliminating d, we get $2c = a + 1$, which tells us that a is odd, and this is impossible by part (a).

Similarly if $p = 1 - i$, then writing $a + i = (1 - i)(c + di)$ tells us that $a = c + d$ and $1 = d - c$. This gives $2c = a - 1$, again implying a is odd, which is impossible.

(d) Since $\text{hcf}(a + i, a - i) = 1$, if $(a + i)(a - i) = b^3$, then both $a + i$ and $a - i$ must be cubes in $\mathbb{Z}[i]$. Thus, we can find $c, d \in \mathbb{Z}$ such that $a + i = (c + di)^3$. Comparing real and imaginary parts, we have

$$a = c^3 - 3cd^2$$
$$1 = 3c^2d - d^3.$$

From this, we have $1 = d(3c^2 - d^2)$, which forces $d = \pm 1$. So then we must have $c = 0$ and therefore $a = 0$. Thus, the only solution to $a^2 + 1 = b^3$ is $a = 0$, $b = 1$.

Index

(a/p) 72
abundant 20
$\alpha(a,p)$ 87
amicable pair 20
arithmetic functions 4
associate 131, 187

binomial coefficient 153

characteristic 124
classification of even perfect numbers 13
commutative ring 112
complex conjugate, $\bar{z}$ 145
complex modulus 189
content 148, 210
coprime 136
cyclotomic polynomials 152

deficient 20
degree 127
descent step 180
discriminant 63
divides 134, 148, 188
Division Algorithm for Euclidean domains 195
Division Algorithm for polynomials 132
divisor 188

Eisenstein's Criterion 151
Euclid's *Elements* 176
Euclidean Algorithm 137, 199
Euclidean domain 195
Euclidean norm 192
Euler's Criterion 66
Euler's ϕ-function 22
    alternative formula for 35
    formula for 33
Euler's Theorem (a generalisation of FLT) 23

factor 134, 148, 188
factor sum formulas for τ and σ 5
Factor Theorem 142
Fermat numbers 19
Fermat's Last Theorem 174
field 122
field of fractions 190
finite fields 124

Gauss's Lemma 76, 149
Gaussian integers 118
gcd 198
greatest common divisors 198

hcf 135, 198
highest common factor 135, 198
hypotenuse 174

important identity for two squares 202
$\mathrm{int}(x)$ 79
integer part of x, $\mathrm{int}(x)$ 79
integral domains 119, 190
irreducible 144, 148, 188

Jacobi symbol, (m/n) 92

Law of Quadratic Reciprocity 83
    an alternative formulation 84
lcm 139, 208
leading coefficient 127
least common multiple 139, 208
Legendre symbol, (a/p) 72
    $(-2/p)$ 80
    $(-1/p)$ 72
    $(2/p)$ 79
    $(3/p)$ 91
LQR 83
    an alternative formulation 84

(m/n) 92
M_p 14
Mersenne numbers 14
Mersenne prime 14
method of infinite descent 180
minimal counter-example 206
monic 131
multiplicative 7
multiplicative functions 7
multiplicative norm 192

norm 192
number-theoretic functions 4

order 25
order of an integer modulo n 25
order of an integer to a composite modulus 25

parity 176
parts 12
parts of a natural number 12
Pell's equation 187
perfect 12
perfect number 12
    even 13
    odd 18
$\phi(n)$ 22
    alternative formula for 35
    formula for 33

polynomial over F 125
polynomial ring over F 125
polynomial ring over $\mathbb{Z}$ 145
prime 188
prime decomposition formula for σ 10
prime decomposition formula for τ 6
primitive 148, 175, 210
primitive root 38, 69
proper factors 134, 148, 188
Pythagoras' Theorem 174
Pythagorean equation 174
Pythagorean triangle 184
Pythagorean triple 175

quadratic character 64
quadratic character of 2, $(2/p)$ 79
quadratic character of 3, $(3/p)$ 91
quadratic congruences 61
quadratic non-residue 64
quadratic residue 64
quotient 132

Rational Root Test 146
reduced set of least positive residues modulo n 22
reduced set of residues modulo n 22
reducible 144
remainder 132
repunits 27
ring 112
ring with 1 113
root 142
roots of unity 152

σ function 4
    factor sum formula for 5
    prime decomposition formula for 10
square-free 11
subfield 124
subring 117
sums of squares 202

τ function 4
    factor sum formula for 5
    prime decomposition formula for 6

UFD 206
unique factorisation domain 206
unit 121, 186

$\bar{z}$ 145
zero divisor 119